Regulators of hemoglobin switching in zebrafish and human models

Regulators of hemoglobin switching in zebrafish and human models

A dissertation presented

by

Jared Jason Ganis

to

The Division of Medical Sciences

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Human Biology and Translational Medicine

Harvard University

Cambridge, Massachusetts

April 2014

Dissertation Advisor: Dr. Leonard I. Zon Jared Jason Ganis

Regulators of hemoglobin switching in zebrafish and human models

Abstract

Hemoglobin switching is a developmental process involving the dynamic transcriptional regulation of multiple *globin* genes. This molecular process involves multiple layer of complexity, and elucidating new mechanisms in this process will result in a more complete understanding of general gene regulation and will likely have direct clinical implications for hemoglobinopathies, such as sickle cell anemia. In this dissertation, I develop and characterize a new model for hemoglobin switching, the zebrafish. I defined and fully annotated the two zebrafish *globin* loci, termed major and minor loci. Both loci contain *α*- and *β-globin* genes oriented in a head-to-head fashion. Characterization of the *globin* expression pattern precisely defined the timing of normal switching and demonstrated that zebrafish, like humans, have two *globin* switches. The locus control region for the major locus was identified and in conjunction with a proximal promoter was able to generate robust, erythroid-specific expression in a transgenic line.

Utilizing the zebrafish system, I performed morpholino knockdown and chemical screens to identify genes and pathways that increase adult *globin* expression in embryos, therefore potentially regulating *globin* switching. The morpholino screen identified the combined knockdown of *tcf7l2* and *ncoa1* as causing significant temporal misexpression, and the chemical screen identified a number of nuclear hormone receptor pathway modulators, including T4, and the Wnt inhibitor XAV939. Further testing demonstrated that the thyroid hormone receptor and

Wnt pathways synergize. The thyroid hormone receptor pathway specifically increases adult *globin* expression at the mRNA and protein levels in both embryos and larva. Conversely, antagonism of this pathway in the adult leads to an increase in larval and embryonic *globins*. Data from K562 and human primary fetal $CD34^+$ cells demonstrate that the regulatory role of the thyroid hormone receptor pathway is conserved in mammals. ChIP-seq analyses in the zebrafish and cell culture systems show that this effect is potentially directly mediated at the *globin* locus. The data presented here demonstrate that both the Wnt and thyroid hormone receptor pathways play a role in the regulation of hemoglobin switching and potential open new area for investigation into clinical treatments for hemoglobinopathies.

Table of Contents

Acknowledgements

This section could not possibly start anywhere besides with my advisor, Len. Without him there would be no lab, idea, project, manuscript, or thesis. His scientific abilities certainly do not need to be detailed here, but the additional effort he puts into mentoring graduate students should not go unmentioned. He knows how much guidance to give, when to give it, and, probably most importantly, how to give it. His ability to be eternally optimistic was particularly essential when the research failed miserably, as it often does. He also takes great care to build the right lab environment to support students in all aspects of graduate school. While I could have done my thesis research elsewhere, I cannot image that I would have learned as much, about science and otherwise.

I would like to thank my Dissertation Advisory Committee for all of their time and valuable input. They were able to guide my project forward and help me avoid the many distractions experimental science can provide. Clifford Tabin's knowledge of development, Barry Paw's knowledge of zebrafish hematology, and Stuart Orkin's encyclopedic knowledge of globin switching and hematology were fantastic resources. I would also like to thank my Thesis Committee, Alan Cantor, George Daley, Christopher H. Lowrey, and Barry Paw for taking the time to read and evaluate my thesis.

This project would also not have left the starting blocks without the help of James Palis. His willingness to share data, review results, and provide input has made him a true collaborator. Numerous others from within and outside of the lab have provided valuable assistance along the way. Eirini Trompouki, Yi Zhou, Anthony DiDiase, and Zi Peng Fan have all made critical contributions to this work. David Wiley has been a great addition to the Zon lab globin switching team for the last few years and been a great person to work with. He not only

continues to make important scientific contributions, but is also able to provide much needed prospective. He has also taken a great deal of time to read and comment on this thesis.

During my time in the lab, I have certainly learned that it takes a lot more to have a successful lab than just great science, and without the help of Dorothy Giarla, Bruce Barut, Christian Lawrence, Isaac Adatto, LiKun Zhang, and Kara Maloney I would not have been able to get anything done in the lab. I have also had the opportunity to work with three great technicians, Rebecca Maher, Elizabeth Riley, and Sruthi Satishchandran, that provided invaluable technical assistance. Working with high school and undergraduate students was another valuable experience in the lab. Particularly, Janelle Lambert was great to work with and in addition to contributing to this work, made significant progress on her own project, which is included as an appendix here.

I cannot thank everyone that has been a part of the Zon lab during my time enough. From answering questions about protocols and jumping in with suggestions to help troubleshoot an experiment, to chatting about anything and everything at lunch and getting a drink after work, they have made the lab a fantastic place to spend six years. I particularly want to thank my fellow graduate students. The lab would not have been the same without you, and would not have been nearly as much fun.

Lastly, I want to thank my parents and Sara. Sara has done a great job of putting up with me while I wrote this thing, and they have all supported me throughout this whole process and, particularly in the case of my parents, well before it.

This thesis is dedicated to

my parents and Sara

Chapter 1

Introduction

Hemoglobin switching is a developmental process involving the dynamic transcriptional regulation of multiple *globin* genes. This process is inextricably linked to normal development, particularly, of course, hematopoietic and erythroid development, which flows from the central role of hemoglobin in the delivery of oxygen to the tissues that are also changing throughout this process. The regulation of this process is fittingly complex, and, while substantial progress has been made in the nearly 150 years since its discovery [1], critical questions still remain. Numerous models have been developed and studied to add to the general understanding of the process, but all have limitations. The study of general development has uncovered numerous genes and pathways controlling the process, some that directly impact *globin* switching and other that have not been implicated. Study of this process is critical not only to gain a better understanding of development but to potentially address diseases associated with hemoglobin and hemoglobin expression.

Erythropoiesis during the Development of Mammals and the Zebrafish

Erythropoiesis, the development of mature functional erythrocytes from precursor cells, is an essential function in vertebrates. This process, however, changes to meet the changing needs of the organisms as it develops [2,3]. During development, erythropoiesis is separated into two major categories, primitive and definitive [4]. The primitive wave is characterized by unipotent progenitors specified from mesoderm, which occurs in the yolk sac in mammals. Erythrocytes and macrophages are produced during this wave in order to oxygenate and help remodel the embryo as it continues to develop [5]. However, this wave is transient, as primitive erythroid progenitors, for example, can first be detected in the mouse at embryonic day (E) 7.25 and become undetectable after E9.0 [4]. As the vertebrate embryo continues to develop, hematopoietic stem cells (HSCs) are specified, and all the mature blood cells produced by these

HSCs compose the definitive wave of hematopoiesis [6]. HSCs emerge in the aorta-gonad-mesonephros (AGM) region, but migrate throughout the developing embryo before homing to their final, adult location in the bone marrow (mammals) or kidney marrow (zebrafish) [7,8].

The red cells produced by these two waves of hematopoiesis, while accomplishing the same general function, are not identical. The size, developmental lineage, gene expression profile, and their sensitivity and response to cytokines vary [4,7–9]. Another, often defining, difference between primitive and definitive red blood cells is their *globin* gene expression profile. In mice, humans, and zebrafish primitive cells express a different set of *globin* genes than adult, definitive red cells [4,7,9].

Within each of these waves, red cells mature from their precursor cells into fully mature erythrocytes through a number of steps (proerythroblasts, basophilic erythroblasts, polychromatic erythroblasts, orthochromatic erythroblasts and reticulocytes) [10–12]. However, the location of these maturational steps (marrow in the adult; bloodstream in the embryo) and whether they are synchronous or not (staggered in the adult; synchronous in the embryo) vary [4,12].

This complex specification and maturation is controlled and executed by a large, complex network of transcription factors, many of which are common to both lineages [4]. GATA1, considered the mater regulator of both primitive and definitive erythroid lineages, is critical for proper specification, maintenance and function [13]. GATA2 is also a key regulator essential for normal erythropoiesis [14,15]. SCL, LMO2, KLF1, and numerous other factors also play a direct role in the successful development of functional erythroid cells, including production of both the protein and heme components of hemoglobin [3,12,13,16–18]. Non-cell autonomous

control of erythropoiesis is mediated by cytokines, the most critical of which is erythropoietin (EPO). The regulatory mechanism of EPO senses the organismal state of hypoxia, and in times of need, induces EPO. This in turn increases the output of erythropoiesis to compensate [19].

Globin expression and regulation in zebrafish and mammals

Structure and function of hemoglobin

Hemoglobin was first discovered in 1840 by F. L. Hünfeld as crystals forming in blood [20], and significant strides in our understanding of this essential protein have been made since. The functional unit of hemoglobin is a tetramer, composed of two α-like and two β-like protein components, known as globins. Each globin carries a prosthetic heme molecule, and each heme group contains an iron molecule, which in turn can combine with an oxygen or carbon dioxide for transport throughout the organism [21]. This process allows the organism to oxygenate tissues not directly in contact with the atmosphere. Cooperative binding by these tetrameric units aids efficient transport of oxygen from oxygen rich environments (lung/gills) to oxygen poor area (body tissues), and conversely for carbon dioxide [21,22]. Globin protein synthesis, as with all proteins, begins with transcription of mRNA from the appropriate genomic locus followed by translation into an amino acid chain. However, because of the tetrameric quaternary structure of hemoglobin and the four prosthetic heme groups, the final functional molecule is only fully functional after amino acids chains generated from two distinct genomic loci and a complex biochemical pathway are combined [23].

Developmental changes in globin expression and the organismal and cellular level

Hemoglobin switching is a developmental process involving the dynamic transcriptional regulation of multiple *globin* genes and occurs in all classes of vertebrates [24]. One reason for

these alterations in *globin* gene expression are the changing oxygen needs of organisms as they develop [25]. Hemoglobin switching can occur at the organismal level, the *globin* composition of the red cell compartment overall changes [4], or at the cellular level, the *globin* genes that an individual cell expresses change over the life of that cell [26]. However, whether the changes in signaling, transcription factor milieu, chromatin structure, or exogenous treatments exhort their effects on different cells or the same cell overtime, they still regulate *globin* gene expression and therefore the *globin* switching process. As mentioned above, switching at the organismal level is often correlated to red cells derived from the primitive wave versus the definitive wave. In both zebrafish and mammals, primitive erythrocytes are derived from mesoderm and transition through a bipotent hemangioblast stage. In the zebrafish, primitive red cells eventually emerge from the region known at the intermediate cell mass (ICM), while in mammalian systems they arise in the yolk sac [4,6,27]. Zebrafish red cells are thought to be the only red cells in circulation for the first 4 days of life [28], but may remain in circulation for more than 7 days [27]. Detection of primitive red cells in mammals is possible as early as embryonic day 7.5 [29]. In both cases the red cells derived from the primitive wave express the embryonic forms of hemoglobin [30,31]. Definitive zebrafish and mammalian red cells differentiate from multipotent progenitors known as hematopoietic stem cells (HSCs), which originate in the aorta-gonad-mesonephros (AGM) region during development. These cells provide a consistent supply of red cell progenitors throughout the life of the organism that express, nearly exclusively, adult *globin* genes [31,32]. Changes at the level of individual cells, termed "maturational" *globin* switching, mainly occur within the primitive wave of mammals as individual cells mature [26]. One or more of the differences between these developmental paths or the signals they receive during them results in these cell types expressing different, specific *globin* genes.

Structure, function, and cis-regulatory elements at the globin loci

The initiation of hemoglobin production begins at a *globin* locus. *Globin* loci are complex regions that contain a number of *cis*-regulatory elements and a number of coding *globin* genes. The general, overarching structure and synteny of *globin* loci has been conserved since the ancestral *globin* locus found in early jawed vertebrates. In humans and mice, there are two *globin* loci, one encoding the *α*-like *globins* and the other the *β*-like. However, this organizational structure is not present in all vertebrates [33], including puffer fish which have two loci, but there is not a strict segregation of the *α*- and *β*-like genes. However, synteny is conserved as genes flanking the human *α-globin* locus are also found upstream of both puffer fish *globin* loci [34]. The number of *globin* genes on each locus also varies between species, including between mice and humans [35]. In humans and mice, these genes are all oriented in the same direction within the loci and organized in order in which they are developmentally expressed [36].

The most critical and influential regulatory elements controlling the expression of these genes is a long-range enhancer known at the locus control region (LCR), which can be over 10kbs upstream of the genes it regulates. It promotes high levels of cell-type specific *globin* gene expression [37] and helps regulates which of the coding *globin* genes is actively transcribed at any one time [38]. The LCR is often characterized by DNase I hypersensitive sites and the binding of critical erythrocyte transcription factors, such as the master regulator GATA1. DNase I hypersensitive sites are formed by several critical erythroid factors including GATA1, EKLF, and NFE2, and the erythroid-specific chromatin structure that develops is essential for proper function of the element *in vivo* [39]. This element does vary in structure in humans and mice and even between the two loci within a species [38,40], but it does retain many conserved functions

across vertebrates as well as some other species [33,41–45]. The LCR interacts directly and physically with the proximal *globin* promoters, creating a "competition" between the promoters for activation by the LCR [38].

The proximal promoters regulating their respective coding genes also play an integral role in regulating *globin* expression. The physical looping of the LCR to interact with discreet proximal promoters forms a structure known as an active chromatin hub [46]. This structure adopts different conformations and composition (i.e. which, if any, *globin* promoters are actively participating) throughout the course of the normal differentiation of erythroid cells as well as when different *globin* genes are being expressed [38,47]. While expression of *globin* genes can occur without the LCR, the LCR enhances expression 100- to 1000 fold over the proximal promoter alone [37,48]. In addition, the LCR cannot act to enhance the expression of the coding *globin* genes without the presence of a proximal promoter to recruit essential *trans*-acting factors. The proper functioning of *globin* loci depends on features of the LCR, proximal promoters, and the precise orchestration of the active chromatin hub they form [36].

While much of the work to define the specific roles of these elements has been completed in model systems where the individual elements, such as individual DNase I hypersensitive sites, can be mutated, deleted, or otherwise disrupted and the effects on *globin* expression observed to infer their specific role(s) [37,38,40,49–52], substantial strides in our understanding of the *globin* locus has originated from work in humans. However, in humans, such experiments can only be conducted through the study of naturally occurring mutations. Both point mutation and large regions of deletion in the *globin* loci have been identified, causing a variety of phenotypes such as β-thalassemia, α-thalassemia, and hereditary persistence of fetal hemoglobin (HPFH). By establishing and correlating the phenotypes and genotypes of these individuals, critical

regulatory regions and transcription factor binding sites have been identified. While it is unfortunate that many of the mutations in the human *globin* loci lead to disease states, identifying and characterizing them has lead to a better understanding of *globin* gene regulation [47,51,53,54].

Regulation of globin switching by trans-acting factors

While the *cis*-regulatory elements found at the *globin* locus are complex, multifunctional elements, they cannot fully explain *globin* switching. As all red cell progenitors transcribe *globin* using the same genomic loci, some extra-genomic differences must be acting to selectively produce particular *globin* genes. The critical role of the *trans*-acting environment was shown through early work in amphibians. It was first shown that the external environment of the cell could influence the *globin* genes expressed by erythroid cells when transplants between two sites of erythropoiesis within an animal produced red cells expressing *globin* genes native to their transplanted location rather than location of origin [55]. The influence of the internal milieu of the cell was demonstrated with the use of heteropolykaryons, by fusing adult *globin* expressing erythroid cells with embryonic *globin* expressing cells from either another frog species or between frogs and mice. The transacting environments stimulated the previously inactive *globin* genes [56]. The reciprocal nature of these experiments, that the embryonic environment could activate embryonic *globins* in adult cells and the adult environment could activate adult *globins* in the embryonic cells, showed the critical influence of the cellular milieu in controlling *globin* gene switching, and the ability to work across species clearly demonstrated the conserved nature of these pathways [56–58]. The *trans*-acting factors, such as transcription factors, act on the *cis*-regulatory elements that compose *globin* loci and modulate their function either directly or indirectly.

The presence of multiple types of hemoglobin in humans has been appreciated and studied for over 80 years [59], resulting in a number of advancements in our understanding of the what factors are involved and their mechanisms of action. The transcription factor BCL11A has emerged as one of the most critical factors in *globin* gene regulation. It was originally identified as a potential regulator because polymorphisms in the gene were associated with HPFH [60–62], and its role established in a number of model systems [63,64]. Sankaran et al. were able to establish BCL11A as a direct, influential repressor of fetal hemoglobin. The full-length isoform of the gene is restricted to expression in the adult, and, when knocked down in a murine erythroid cell line or differentiated human primary adult $CD34^+$ cells, causes an increase in fetal globin expression. ChIP-seq analysis showed the protein binding directly to key DNase I hypersensitive sites in the LCR and other regions within the *globin* gene cluster itself [63,65]. Following investigations demonstrated that differential regulation of BCL11A between mice and humans is major factor contributing to the divergent *globin* switching patterns observed between the two species [64]. As a critical repressor of fetal hemoglobin in the adult state, it was hypothesized that knockout of this gene in erythroid cells would sufficiently increase HbF levels to correct the hematologic defects associated with sickle cell disease. This result was confirmed using an erythroid specific knockout system in a humanized mouse model of sickle cell disease [66]. However, despite the substantial strides in understanding the role of BCL11A and the BCL11A pathway in *globin* switching, there are still significant gaps in our knowledge of the mechanisms surrounding the protein's function. For example, the regulation of *BCL11A* itself in the erythroid lineage needs to be more clearly defined. Naturally occurring human variations in transcription factors binding sites within the gene body have been shown to affect *BCL11A* expression, and therefore *globin* expression [67]. However, many of the upstream regulators of

BCL11A remain to be elucidated, including links to signals received from the external environment that can regulate *globin* expression [55]. Downstream of the regulation of *BCL11A* expression, further investigation is still needed to determine the exact mechanism of its regulation of the *globin* locus, which can help to further clinical efforts to intervene in the pathway [68].

Some genes that interact with BCL11A have been determined, and they also contribute to the regulation of the *globin* genes. SOX6 is a HMG-box-containing transcription factor that cooperates with BCL11A to repress γ-globin. The factors co-occupy key sites in the LCR, proximal promoters, and other regions of the globin locus as well as physically interact. These factors work together to modify the overall structure of the looping between the LCR and the proximal globin gene promoters, which in turn contributes to alterations in globin gene expression patterns [65]. Other factors that physically interact with BCL11A have been identified include TRIM28, NuRD complex members, MARTIN3, SWI/SNF complex members, and DNMT1. However, the mechanism of action of these factors remains unclear as well as how, if at all, they affect the regulation of globin switching [63,68,69].

Other factors such as EKLF [70,71], LDB1 [72,73], CTCF [74,75], MYB [76], and IKAROS [77,78] have been demonstrated to have or implicated in a role in the regulation of hemoglobin switching, in addition to many others [38,40,47,68,79]. The mechanisms of many of these factors and how they integrate into the mechanism of central factors, such as BCL11A, and the general globin regulation network remain extremely unclear. It is also likely others factors that contribute to globin regulation and switching will continue to emerge as new techniques and models are used to investigate the system.

Current and Potential Models in the Study of *globin* Regulation

Current models of globin switching

A number of models are currently used to investigate the process of *globin* switching. The models vary in ease of use, techniques that can be applied to them, base line *globin* expression patterns and transcriptional milieu, normal behavior, validity of particular types of inferences, and relevance to clinical use in humans. In combination, they offer a wide breath of options, though no one system is ideal to a particular experiment and certain combinations of features remain unfulfilled.

Cell culture models include immortalized cell lines as well as primary erythroid progenitors that are differentiated in culture. Immortalized cell lines such as the human erythroleukemia line K562 [80–82], the murine erythroleukemia line GM979 [83,84], and other murine erythroleukemia (MEL) lines [85] have been used extensively to study to globin gene expression and regulation. These cell lines offer the ability to generate large amounts of material to perform screens, ChIP-seq experiments, protein mass spectrometry and other larger scale experiments. They also afford the opportunity to utilize an isolated system, free from external influences outside of the cell type in a mammalian context, and a variety of globin expression profiles including mainly human γ-globin in K562 cells [81] and murine embryonic and adult globin in GM979 cells [83]. However, they do not accurately recapitulated any endogenous cell type and may have a number of pathways involved in any given process disrupted due to their immortalization and high number of passages. Primary cells afford similar benefits as cell lines, but are more limited in number and ease of acquisition. They also provide a system more closely resembling the endogenous state. A particularly essential primary cell tool in the study of

hemoglobin switching are primary human $CD34^+$ cells [86–89]. These cells can be derived from adult peripheral blood, cord blood, and fetal liver, and when differentiated express the appropriate globins for their stage, adult, a mixture, and fetal, respectively [90]. This provides a system where endogenous differences, as well as induced differences, can be studies in a human system, in a relatively large number of cells.

Whole model organisms are also often used in the study of hemoglobin switching. Organisms such as frogs [91,92], salamanders [93,94], and mice [95–97] have been utilized for decades. These models allow investigators to probe effects and changes that occur at the level of the whole organism, in the context of an environment with multiple cell types, and throughout the natural developmental process. In many cases there are practical limitations to utilizing whole organisms for particular experimental designs, such as high throughput screens. These whole organismal systems, while generally conserved with humans, may vary in key elements under investigation. For example, the murine globin switching process, while in a conserved mammalian system, has significant differences when compared to humans [64,98]. However, efforts have been made to engineer systems with different combinations of attributes and advantages. Mice transduced with genomic material from humans have been used extensively to study and discover key features of the *cis*-regulatory elements surrounds the globin loci [37,99,100]. Stable transgenic mice containing the entire human β-globin locus have been developed [52] and successfully studied [101], as well as disease state mice, such as sickle cell models [66,102,103]. These animals can be used to test alterations to human regulatory systems and disease states, but in the context of a whole organism, allowing for a more holistic understanding of the human system.

From the initial discovery of hemoglobin [20], the presence of multiple forms of hemoglobin in human blood [1,59], sickle cell disease [104], and thalassemia [105,106], studying humans directly has also provided substantial contributions to our understanding of hemoglobin and hemoglobin switching. Specific mutations such as the Corfu mutation, a naturally occurring 7.2 kb deletion in the human β-globin locus, has helped to define a regulatory region that influences the expression of both the fetal *γ*-globin and adult β-globin. The region was also later shown to contain an important BCL11A binding site, potentially explaining part of the regions function [68,107]. The ability to directly examine the role of mutations and polymorphisms continues to contribute to the field, such as through genome wide association studies [60–62]. These naturally occurring mutations play an essential role in helping to understand globin switching [68].

The zebrafish as a model for hemoglobin switching

The zebrafish model, while used extensively in hematology research [108], has been under utilized in the study of hemoglobin switching. Between the embryo and adult systems of the zebrafish, it offers a unique combination of features that could help to advance our understanding of the process. One of the most significant advantages of the zebrafish system is the ability to perform large scale screens on a whole organism. The ability to generate a large number of optically clear, externally fertilized embryos quickly, and then easily assay temporal and spatial gene expression by *in situ* hybridization allows both chemical and knockdown screens to be performed in a high-throughput manner [109–112]. It is also possible, while in lower throughput manner, to utilize the adult zebrafish for analysis [113]. These features are critical to extend the examination of *globin* switching beyond previously identified factors, and

the ability to easily access and utilize animals at all stages of development is particularly essential.

Despite these advantages, there are also a number of limitations to screening in the system. When performing knockdown screens, the mRNAs of genes are typically individually targeted using morpholinos (GeneTools, Philomath, OR) designed using published, publically available genomic data. However, due to imperfections in the zebrafish genome assembly, selection of well defined transcription start sites and splice sites, which are needed for morpholinos targeting, is not always possible. In addition, the whole genome duplication that took place in the zebrafish evolutionary lineage [114] has resulted in multiple copies of genes only represented by a single copy in mammalian systems. It is often not possible to determine prior to screening if the function of the mammalian protein is preserved in one, both, or neither of the zebrafish genes. Lastly, not all designed morpholinos will effectively target and knockdown their intended transcript, creating a false negative rate of approximately 11% [110]. Chemical screens are also limited by a number of factors. A number of the chemicals tested will most likely have unknown or multiple molecular targets or cause developmental delays, which can complicate downstream analysis [112]. Even chemicals for which the target is defined in mammalian systems may have a less potent, off-target, or an otherwise substantially different effect in zebrafish due to the difference in protein sequence, metabolic processing, or bioavailability. The permeability of the zebrafish chorion, which is not always removed prior to screening, also may alter the effects of certain chemicals in the library [115]. Finally, while there is substantial conservation of gene and pathway function between zebrafish and mammals, differences may exist that render some information gleaned from zebrafish studies incorrect in the mammalian system.

As hemoglobin switching is inherently a process of development, the ability to easily compare, contrast, and interrogate all stages is central to gaining a better understanding of the process. The zebrafish offers easier access to, and therefore easier ability to manipulate, whole organisms from the single cell stage through adulthood on an accelerated time scale. The ability to easily isolate peripheral blood from nearly all stages of development allows the direct study of these primary cells in relatively large numbers and all stages of interest. However, due to their smaller sized compared to mice, tissue collection from individuals is limited, but can be compensated for with the use of additional animals.

The zebrafish is a particularly advantageous model for studying hematology, including hemoglobin switching. The similarity to the human hematopoietic system and its regulation [3,7,116] is crucial to making observations that are relevant to the human system quickly. The availability of established genetic mutants that have hematological phenotypes [28,116–119] allows follow up work to better characterize genes, chemicals, pathways identified by experimenting with them in known hematological contexts (e.g. no primitive red cells). In addition, red cells are expendable in zebrafish embryos and early larva, making the study of genes that significantly disrupt red cells in early development more tractable [108,120,121]. In conjunction with these mutants, there is also an extensive set of transgenic fish for the study of hematology, such as the GATA-1/GFP line [122], and additional lines can be readily developed [123].

Current State of *globin* Research in the Zebrafish Model

Globin expression throughout the development of the zebrafish

As noted earlier, red cells in the zebrafish are generated in both the primitive and definitive waves. The primitive wave is born in the ICM [27], is the only circulating red cell population for the first 4 dpf [28], and is can be detected past 7 dpf [27]. Definitive red cells differentiate, through a series of steps, from hematopoietic stem cells and begin to contribute to circulation at around 3.5 dpf [27]. The initial onset of *globin* expression begins at approximately the 12 somite stage [30] from at least one of the two known *globin* loci [124]. From analysis of cDNA libraries generated from 54-56 hpf wild-type zebrafish embryos, six *globin* genes were identified and termed α_{e1}, α_{e2}, α_{e3}, β_{e1}, β_{e2}, and β_{e3} [30], one of which, β_{e1}, had previously been identified in a cDNA library generated from 3-35 hpf embryos [125]. By protein analysis at this stage, only three globin proteins were identified (α_{e1}, α_{e3}, and β_{e1}), but expression of mRNA at 24 hpf, 48 hpf, and 120 hpf was confirmed by *in situ* hybridization for α_{e1}/α_{e2} (sequence similarly prohibited separation), α_{e3}, β_{e1}, and β_{e2}, while β_{e3} expression was not examined at 120 hpf but confirmed at the 2 previous stages. The expression of all identified embryonic *globins* was substantially reduced at 72 hpf for unknown, potentially technical, reasons [30]. However, the discovery effort here, cDNA library analysis at 54-56 hpf, is not comprehensive. Additional *globin* genes with similar sequences could be expressed off of different genomic loci, some *globin* genes could be expressed in the embryo or larva outside of this narrow time window, and particular clones expressed at low levels could be overlooked using this method. When genomic fragments in phage and bacterial artificial chromosomes were sequenced and annotated with the known *globin* sequences, it was discovered that the embryonic *globin* sequences were arranged in a tail-to-tail fashion and linked with known adult *globin* sequences. This demonstrated that, while there are two zebrafish *globin* loci, they are not arranged in a similar manner to the mammalian loci: (1) they are not segregated into *α*- and *β-globin* loci, and (2) they are not

organized in the same direction in the order to expression. The results are consistent with other known fish species and the expected evolution of the *globin* loci [33,34,126]. However, these methods are limited in fully characterizing and completely annotating the two known zebrafish *globin* loci.

Globin expression in the adult zebrafish has also been analyzed. Utilizing cDNA libraries generated from adult peripheral red cell, three *globins* were identified and termed α_{a1}, β_{a1}, and β_{a2}. At the protein level, HPLC identified five expressed *globin* genes, those predicted to be transcribed from the identified cDNAs, α_{a1}, β_{a1}, and β_{a2}, and the newly termed α_{a2} and α_{e3}. Work identifying the genomic location of these genes demonstrated a head-to-head ordination of an α- and β-*globin* gene, demonstrating a non-mammalian system of organization [32]. The discrepancy between the number of *globin* genes expressed by analysis of the cDNA libraries and the protein products could be the results of differences between zebrafish, limitations of cDNA libraries, post-transcriptional modifications made to the globins, changes in *globin* expression over time (i.e. the cDNA for some or all of the proteins observed were previously transcribed and have been degraded), or technical issues resulted in aberrant identification of additional globin proteins (e.g. some of the proteins identified represent other red cell proteins). Taken together, these observations demonstrate that the zebrafish has at least one *globin* switch between an embryonic stage, defined by the expression of six *globins* and occurring at least between 54-56 hpf, and an adult stage, defined by the expression of between two and five *globins* [30,32].

Later work expanded this analysis by analyzing *globin* expression at 24 different developmental stages between the one-cell stages and sexual maturity at 54 dpf using RT-PCR [127]. The *globins* were grouped into five categories, embryonic α and β, adult α and β, and a

newly identified embryonic *α-globin* termed *hbαX*, limiting the analysis in describing the fluctuations between the distinct *α-* and *β-globins*. *HbαX* was identified by analysis of zebrafish genomic assemblies and comparing sequence homology to other identified *globins* within and from outside the species. Its embryonic nature was determined by the timing of expression (see below). Analysis of one-cell stage embryos showed high levels of expression of all five *globin* categories as compared to the next developmental stage examined (4 hpf), implying maternal contribution of a wide variety of *globin* mRNAs [127]. Zygotic embryonic *globin* expression then begins to be detected around 12-18 hpf [127], consistent with previous work [30], and peaks at approximately 72 hpf [127]. Expression of all three embryonic *globin* groups continued through the next 14 developmental stages and into adulthood (54 dpf). A significant increase in the expression of the adult *globin* groups was first detected at 14 dpf, with continued expression through the remaining 5 developmental stages tested and into the adult stage [127]. These results do not agree with earlier observations where only adult *globin* mRNAs are detected in the adult stage [32], but, if accurate, could potentially explain some of the inconsistencies between the number of *globins* observed in the mRNA and protein analyses. Within both the embryonic and adult *globin* expression ranges, the expressions levels also fluctuate substantially. The expression of both the adult *α-* and *β-globin* increase around 3-4 dpf before decreasing again at 5 dpf, and the expression of all three embryonic *globin* groups significantly decrease between 7-9 dpf before recovering to previous levels [127]. These fluctuations, in contrast to observing a defined period of expression proceeded by a period of increase and followed by a period of decrease, may indicate that different *globins* within the groups are not adhering to the same expression pattern. This indicates that grouping the zebrafish *globin* genes into five categories may not provide sufficient resolution to accurately assess the pattern of *globin* switching in the zebrafish.

Known regulators of globin expression in the zebrafish

While the *globin* genomic loci and normal developmental switching pattern remain incompletely characterized, environmental and transacting factors that regulate the process have begun to be identified. One of the environmental factors that triggers a shift in the expression of *globin* genes is hypoxia. As *globin* genes are integrally involved in the oxygen state of an organism, it is perhaps not unexpected that one of the physiological responses to a reduced oxygen state is to modulate *globin* expression. When exposed to hypoxic conditions, zebrafish *globin* mRNA, α and β, expression decreases [128]. This is in contrast to observations in mammals and other fish species, such as medaka [128,129]. While it is possible that this may be a response unique to zebrafish, it is also possible that the limited *globin* genes observed, an unknown collection of *globins* grouped and assessed only as α and β, do not reflect total *globin* levels. In humans, hypoxia can induce higher levels of fetal *globin* [130], so it is possible a similar mechanism is increasing the level of *globin* genes that were not assessed, such as larval *globins*, in addition to decreasing the levels of the *globins* that were tracked. Some of the downstream factors of this environmental trigger of fetal *globin* induction are known, such as HIF-1 [131], but the mechanisms of action of these factors in zebrafish hemoglobin switching are largely undetermined.

As mentioned above, the KLF genes play a critical role in *globin* expression in mammals [70,71], and the family has also been implicated in the regulation of *globin* expression in the zebrafish. *Klf4* has been shown to regulate *globin* expression only in primitive erythrocytes, which express embryonic *globins*, and not definitive erythrocytes, which express adult. Knockdown experiments demonstrate that the gene is essential for embryonic *globin* production, but dispensable for adult [132]. Similarly, other zebrafish *klf* family members have been shown

to have distinct roles in embryonic *globin* expressing primitive cells and adult *globin* expressing definitive cells [133,134]. In addition, some regulatory mechanisms governing differential *globin* expression likely act though *klf* family members. *MiR-144* has been shown to negatively regulate embryonic *α-globin*, but not *β-globin*, though its regulation of *klfd*, which binds exclusively to the promoter of α_{e1} and not β_{e1} [135]. Despite the high levels of expression, layers of regulatory controls, and other cellular resources dedicated to the production of primitive zebrafish red cells, they, as mentioned earlier, are not essential to life in the developing embryo for approximately 7 days [108,120,121].

Critical Pathways that Regulate Development in Zebrafish and Mammals

The role of the Wnt pathway in development

The Wnt pathway is one of the most critical, diverse, and complex pathways regulating development and adult homeostasis. There are three branches of the pathway: (1) the canonical pathway, which is converges on β-catenin, (2) the non-canonical pathway, and (3) the Wnt/Ca^{2+} pathway, both of which are β-catenin independent [136]. The pathway has known roles in early embryonic patterning and axis formation, cell fate decisions, organ development, appendage patterning, and disease states, namely cancers [136–141]. The Wnt pathway has been shown to play a particularly key role in hematopoiesis, playing a role in the self-renewal of HSCs, proliferation, the cell fate decisions within the lineage, and during the differentiation of a number of cell types [140,142–144]. The Wnt pathway is a highly conserved pathway from *C. elegans* to humans, including zebrafish [136,145]. Briefly, in canonical Wnt signaling, the central factor is β-catenin. In the inactive state, it is degraded by the destruction complex (composed of AXIN, APC, GSK3β, and others) in the cytoplasm, preventing it from entering the nucleus. In the

nucleus, the protein Groucho binds to TCF family proteins and modulates gene expression. In the active state, Wnt ligand stimulates the receptor complex of Frizzled and LRP 5/6 proteins, which triggers the sequestration of critical destruction complex proteins. In the absence of the destruction complex, β-catenin accumulates, enters the nucleus, binds to TCF family members and modulates gene transcription [136]. Despite the numerous and critical role the diverse arms of the Wnt pathway play in development and hematopoiesis, there is no known direct role of the pathway in globin switching.

The role of the thyroid hormone receptor pathway in development

The nuclear hormone receptor pathway is another critical and well conserved developmental and homeostasis pathway. It is composed of over 30 individual receptors in humans, only a fraction of which have identified endogenous ligands. Together, they are known to play roles in embryonic development, metabolism, cell identity, and diseases related to these functions such as cancer and obesity. In addition to the receptors and ligands themselves, the nuclear hormone receptor family contains a number of coregulators and heterodimeric receptor partners. The coregulators are grouped into the corepressors, such as NCOR1 and NCOR2, and coactivators, such as NCOA1 and NCOA2, and the RXR proteins are utilized by a number of the other nuclear hormone receptors as heterodimeric partners, only binding to DNA with an RXR partner [146]. The thyroid hormone receptor pathway is a member of this superfamily that plays a particularly central role in early development and metabolism, as well as diseases such as diabetes [147,148]. The ligand for the thyroid hormone receptors (TRs) is thyroid hormone, 3,3',5-triiodo-L-thyronine (T3), which is generated from the partially active precursor L-thyroxine (T4) and processed to T3 in mammals by either of two deiodinase enzymes, D1 or D2. T4 can also be inactivated to reverse T3, and both T3 and reverse T3 can be degraded further to

the inactive T2. These processing steps allow for finer control of thyroid hormone activity as the correct balance of both precursor hormone and deiodinase enzymes must be preset to generate T3 [149]. Under normal cellular conditions in which a TR is expressed, the receptor is bound to thyroid hormone response elements (TREs) with an RXR partner. The presence or absence of ligand regulates which other proteins are bound in the complex. In the presence of a saturating amount of ligand, they are bound to a coactivator, and in the absence of ligand to a corepressor, while an intermediate amount results in a dynamic equilibrium between the two states. While the nomenclature may indicate that gene transcription is activated in one state and repressed in another, the effect is not correlated to the coregulator or ligand state, but whether the TRE is a "positive," active in the present of ligand, or "negative," active in the absence of ligand, target [150]. However, emerging evidence also suggests that variations in the TRE sequence can also determine whether the TR needs an RXR partner or coregulator to modulate gene expression [151]. In humans and mice, there are two distinct TR receptors, TRα and TRβ, which have overlapping but separate roles and can be expressed in different tissues at different times. In zebrafish, there are three receptors, *thraa*, *thrab*, and *thrb* [152]. The thyroid hormone regulates, and is essential to, metamorphosis in frogs and other amphibians [153,154] as well as teleost, including zebrafish [155], and T3 and T4 levels are modulated accordingly [156]. In frogs, a direct link between the thyroid hormone receptor pathway and globin switching has been established as an element of metamorphosis [157,158]. While mammals do not undergo classic metamorphosis as amphibians and teleosts do, thyroid hormone has been linked to like postembryonic developmental changes reminiscent of metamorphosis [159]. The two thyroid hormone receptors are known to have distinct functions in mammals from the study of a large number of mouse mutants, and no known role in globin switching [160–162], but hematopoietic

and erythroid defects have been identified, mostly associated with TRα [163,164]. The thyroid hormone receptor pathway is also known to interact with other essential developmental pathway, such as the Wnt pathway [165–168], but this intersection has not been observed to play a role in hemoglobin switching.

Thesis Review

This thesis fully characterizes the zebrafish as a model for hemoglobin switching and successfully utilizes it to identify and characterize pathways involved in *globin* gene regulation in the zebrafish and mammals, which could be modulated to treat hemoglobinopathies. By fully annotating both *globin* gene loci and establishing the normal *globin* expression pattern for each *globin* from the onset of expression into adulthood, disruptions and modulations to this pattern could be observed. Through this process, two *globin* switches were identified, analogous to the pattern observed in humans. In addition, the mapping of DNase I hypersensitive regions and the binding of the erythroid master regulator Gata1 were used to locate the major locus LCR. This region was later found to enhance robust, erythroid-specific reporter expression in a transgenic zebrafish line.

Through morpholino and chemical screens, the Wnt and thyroid hormone receptor pathways were identified as regulators of the normal *globin* switching process as they disrupted the baseline expression of the adult *globin* gene α_{a1} in the embryo. These pathways are shown to cooperate in the regulation of *globin* genes through co-treatments with both morpholinos and small molecules. Further analysis demonstrates that the thyroid hormone receptor pathway upregulates both adult *globin* mRNA and protein, with, at the doses used, little effect on other aspects of red cell and general development. Investigating this pathway in the adult shows that

the pathway is not only used in the establishment of the adult states, but also in its maintenance as the treatment of adult with a thyroid hormone receptor antagonist induces embryonic *globin* mRNA and protein. The function of this pathway is conserved in the mammalian system, as demonstrated through K562 and human $CD34^+$ cell culture. ChIP-seq experiment in both the zebrafish and human systems show that this pathway works directly on *globin* loci to modulate gene expression. By establishing the zebrafish as a model of hemoglobin switching and identifying and characterizing two pathways that regulate *globin* expression, additional avenues for research into the regulation of *globin* switching and potential therapies for hemoglobinopathies have been opened.

Chapter 2

Zebrafish *globin* switching occurs in two developmental stages and is controlled by the LCR

Attributions

This manuscript was published in *Developmental Biology* (2012 Jun; 366(2):184-94)[169]. I wrote the manuscript, prepared all figures for publication, and incorporated all necessary revisions. I verified the genomic alignments and annotations of the major and minor zebrafish globin loci. I also cloned and subcloned additional globin genes for the preparation of labeled RNA probe for the completion of the *in situ* hybridization experiments, which were completed under my supervision by Janelle Lambert, a Harvard undergraduate. I identified and subcloned the DNase I hypersensitive and Gata1 occupied regions of the LCR for testing and completed the testing of the enhancers *in vivo*.

Nelson Hsia was the second primary author. He generated the initial genomic and protein sequence alignments and developed the RT-PCR primers for, and carried out, the RT-PCR experiments. He also cloned the full length LCR and promoter from the zebrafish genome, and contributed to the manuscript.

Eirini Trompouki carried out the Gata1 ChIP-seq experiment and Zhiying Jia carried out the H3K4me3 ChIP-seq, both with bioinformatic assistance from Anthony DiDiase. Peter Sabo, Molly Weaver, and Richard Sandstrom completed the DNas I hypersensitivity analysis designed by John Stamtoyannopoulos with assistance from the University of Washington High-Throughput Genomic Center staff. Jill de Jong completed the FACS and FACS analysis of the LCR-GFP reported line. Yi Zhou assisted in assembling and annotating the genomic loci. Owen Tamplin provided the pME-MinPro-GFP, and Hsuan-Ting Huang helped review the manuscript.

Abstract

Globin gene switching is a complex, highly regulated process allowing expression of distinct *globin* genes at specific developmental stages. Here, for the first time, we have characterized all of the zebrafish *globins* based on the completed genomic sequence. Two distinct chromosomal loci, termed major (chromosome 3) and minor (chromosome 12), harbor the *globin* genes containing *α/*β pairs in a 5'-3' to 3'-5' orientation. Both these loci share synteny with the mammalian *α-globin* locus. Zebrafish *globin* expression was assayed during development and demonstrated two *globin* switches, similar to human development. A conserved regulatory element, the locus control region (LCR), was revealed by analyzing DNase I hypersensitive sites, H3K4 trimethylation marks and GATA1 binding sites. Surprisingly, the position of these sites with relation to the globin genes is evolutionarily conserved, despite a lack of overall sequence conservation. Motifs within the zebrafish LCR include CACCC, GATA, and NFE2 sites, suggesting functional interactions with known transcription factors but not the same LCR architecture. Functional homology to the mammalian *α-LCR* MCS-R2 region was confirmed by robust and specific reporter expression in erythrocytes of transgenic zebrafish. Our studies provide a comprehensive characterization of the zebrafish globin loci and clarify the regulation of *globin* switching.

Introduction

In organisms dependent on functional hemoglobin for oxygen transport, regulation of its production is vital, and misregulation can have catastrophic effects. An elaborate regulatory mechanism has evolved to govern *globin* production and *globin* switching, the process by which precise changes occur in α- and β-*globin* production in an organism throughout development. *globin* switching occurs in many species across ontogeny, including humans and other mammals,

and relies on both highly conserved and unique *cis*- and *trans*-regulatory elements [33,38,40,43,170,171]. A driving force of *globin* switching at the organismal level is a series of waves of hematopoiesis defined by the production of erythrocyte precursors in different anatomical locations [4] "Maturational" *globin* switching, or the switch in *globin* production of an individual cell as it matures through erythropoiesis, also plays a critical role in defining the *globin* expression signature of the organism as a whole [26].

Genetic disorders in which this process is disrupted, either through the mutation of regulatory regions or *globin* coding sequences themselves, are collectively known as hemoglobinopathies [38,98,172]. These disorders include the thalassemias and sickle cell disease [173–175], which remain a major health concern worldwide [176,177]. The molecular complexity and clinical relevance of *globin* switching has made it an area of intense basic and clinical research; sickle cell anemia was the first disease for which the molecular basis was described [176,178,179].

Fundamental mechanisms of gene regulation [96,180], particularly those of long-range regulatory elements, were initially discovered and studied in this system [37,170]. Clinical observations, including that higher levels of persistent fetal Globin in sickle cell patients ameliorates symptoms, have focused further research towards influencing this *globin* switch as a treatment option [176,181]. Despite these substantial efforts and even breakthroughs, aspects of the mechanism of *globin* regulation remain unclear and a cure for the hemoglobinopathies elusive [176].

Conserved long-range regulatory elements play a critical role in *globin* expression and switching. These enhancer regions are typically, as in the *globin* locus, characterized by DNase I

hypersensitive sites (HS) annotated by their position (in kilobases) upstream of the *globin* coding sequences [170]. The DNase I hypersensitive site ~40kb upstream (HS-40) of the α-*globin* locus in humans has been demonstrated to be essential for proper *globin* expression, with other hypersensitive sites spanning as much as 150kb playing roles in the process [35,38,182]. The conservation in coding, noncoding regulatory and overall synteny of the *globin* loci can be traced to the ancestral *globin* locus present in early jawed vertebrates, which contained both *α*- and *β*-*globins*. The locus has diverged over time and segregated into separate α and β loci after the divergence of amphibians [33,34], but essential functions have been shown to be conserved [33,34,41–43,45,183]. Analysis of this conservation in noncoding regions reveals the presence of multispecies conserved sequences (MCS), which when aligned with DNase I HS often define a functional regulatory sequence. This has allowed for the definition of homologous regions in different species, as well as guided identification of new regulatory elements [40].

The zebrafish, *Danio rerio* has already been established as an important model to study developmental hematopoiesis [184,185]. Advantages include high fecundity, accelerated development, external fertilization, and transparent embryos, allowing for the real-time *in vivo* visualization. Importantly, the embryos do not require hemoglobin or red blood cells through at least the first 15 days of development [186], allowing for detailed loss-of-function studies not possible in other organisms. Similarly, these features allow for large-scale chemical screening not feasible in mammals [187]. Of particular interest is the ability to generate transgenic organisms quickly and perform whole-organism, live fluorescent imaging.

The adult [32] and embryonic [30] *globins* in the zebrafish have been characterized by identifying some of the adult and embryonic *globins*, both *α* and β, within both *globin* loci. Here, the first detailed elucidation of the *globin* expression pattern changes allowed the

observation and characterization of embryonic-to-larval and larval-to-adult *globin* switches during development, consistent with humans [98]. A conserved regulatory element has also been revealed by DNase I hypersensitivity mapping and GATA1 binding data, which we show functionally drives robust and specific expression in red blood cells, validating these techniques for identifying regulatory regions. Additional putative enhancer elements in the major locus and minor locus were also observed and warrant further investigation. These data fully characterize the zebrafish as a model of *globin* switching by defining the coding regions, synteny, genomic structure, regulatory regions and expression pattern of the zebrafish *globin* gene loci as well as demonstrate the level of conservation at both the molecular and functional level between zebrafish and humans.

Materials and Methods

Zebrafish maintenance

Zebrafish were staged, raised, and maintained as described [188,189]. All zebrafish experiments and procedures were performed as approved by the Children's Hospital Boston institutional Animal Care and Use Committee.

RNA isolation and cDNA preparation

Pools of fifty zebrafish from the 18 somite stage (ss) through 32 days postfertilization (dpf) were collected and homogenized in Trizol reagent (Invitrogen, Carlsbad, CA). Small pools of adult fish were also homogenized in Trizol, and total RNA was isolated using the manufacturer's protocol and further purified with the RNeasy kit (Qiagen, Germantown, MD). Subsequently, 1 μg of total RNA was used to generate cDNA using the Superscript III system (Invitrogen).

Quantitative real-time PCR

Quantitative real-time PCR was performed on an iCycler IQ5 Real-Time PCR detection system (Bio-Rad, Hercules, CA) using SYBR Green Supermix (Bio-Rad), and fold change was determined using the Gene Expression Analysis for iCycler iQ Macro (Bio-Rad).

Synthesized cDNA was diluted 1:2, and 1 μl was used in a 20 μl SYBR green PCR reaction (SYBR Green Supermix, Bio-Rad) containing 200 nM of forward and reverse primers. Beacon Designer primer software was used to design qPCR primers to unique zebrafish *globin* genes. Careful attention was made to avoid haplotype-specific regions, and optimized primer pairs are shown in Table 2-1. The products were amplified using a two-step program cycling at 95°C for 10 sec and 57°C for 30 sec followed by melting curve analysis. Gene expression levels were normalized to an erythrocyte-specific membrane gene, band3, and PCR efficiency curves were done for all primer pair sets to assign absolute values to each gene. Data were plotted in two separate groups of α- and β-*globins* as a percentage of each *globin* during each time point during the developmental profile.

Generation of zebrafish transgenics

The LCR and α/β adult *globin* 2 (α/β_{a2}) bidirectional promoter were, respectively, isolated by restriction enzyme and PCR amplified and cloned into pEGFP-1 vector to create the α/β_{a2}-GFP construct. Additional LCR fragments were amplified by PCR and assembled into expression vectors containing transposon elements using the Gateway system (Invitrogen).

A 1.2 kb fragment of the α/β adult *globin* 2 (α/β_{a2}) bidirectional promoter was PCR amplified from the zC113F11 bacterial artificial chromosome (BAC) clone and ligated into the

Table 2-1: Sequences of primers used in quantitative real-time PCR.

Gene	PCR primer	Sequence 5'→ 3'
$\alpha_{e1a/b/c}$	forward	CCAGGATGTTGATTGTCTAC
$\alpha_{e1a/b/c}$	reverse	CAGTCTTGCCGTGTTTC
α_{e3}	forward	CCTAAGCCCCAACTCTC
α_{e3}	reverse	CTCCCTTCAGGTCATCC
α_{e4}	forward	GTGGCGGAAGAAATTGG
α_{e4}	reverse	ATGGAACCTGCTAAGTGG
α_{e5}	forward	CCTCAGACCAAGACCTAC
α_{e5}	reverse	TCAGACAGAGCCAAAGC
$\alpha_{a1a/b}$	forward	CAAGGCTGTTGTTAAGGC
$\alpha_{a1a/b}$	reverse	ATTCTGGCGAGGGCTTC
α_{a2}	forward	GCGAATGCTTACTGTCTAC
α_{a2}	reverse	TGTCACGAGGATGTTATGG
$\beta_{e1a/b/c}$	forward	CTTGACCATCGTTGTTG
$\beta_{e1a/b/c}$	reverse	GATGAATTTCTGGAAAGC
β_{e2}	forward	ACTATGAGGAGGCTGGAC
β_{e2}	reverse	CGGCGTAGGTGTTCTTG
β_{e3}	forward	ATGCTTGGTCGTCTATCC
β_{e3}	reverse	ATGATTGCCTCTGTGTTG
$\beta_{a1a/b}$	forward	ATATTTCGCCACATTCG
$\beta_{a1a/b}$	reverse	GCATAGGTGTTCTTGA
β_{a2}	forward	ATTGCGAGTGTCTGGAG
β_{a2}	reverse	CCGTGTTCTGAAACTTTGG
slc4a1	forward	AGTTAGAGCCTGTTGTATCC
slc4a1	reverse	TCCATCACCACCGAATCC

pEGFP-1 vector to create the α/β_{a2}-GFP construct. A restriction fragment cloning approach was undertaken to excise the intron containing the conserved LCR.

The zC113F11 BAC was digested with Kpn I/ Sma I to liberate a 6.2 kb fragment containing intron 5 and the adjacent exons. This fragment was ligated upstream of the α/β_{a2} promoter to complete the α-LCR-α/β_{a2}-GFP transgenic construct. This construct was liberated from the pEGFP-1 vector by digestion with *Kpn I* and *Nae I* and gel purified using the QIAEX II gel extraction kit (Qiagen). GFP positive LCR-α/β_{a2}-GFP were raised and outcrossed to identify germline founders.

Additional reporter constructs were generated by PCR amplifying the desired portions of the LCR-α/β_{a2}-GFP transgenic construct into the Gateway entry pENTR 5'-TOPO vector (Invitrogen), into which the α/β_{a2} promoter was sub-cloned using the restriction enzyme BamHI downstream of the LCR fragment in the cases where it is included in the final expression vector. The appropriate 5' vector, vector #383, vector # 302 and vector #394 (Kwan et al., 2007) were assembled into the final expression vectors DNaseIpeak-LCR-α/β_{a2}-GFP and GATA1peak-LCR-α/β_{a2}-GFP using LR Clonase II Plus (Invitrogen). The expression constructs containing the enhancer elements and the mouse minimal promoter were similarly assembled utilizing the Gateway system (Invitrogen). The respective enhancer fragments were cloned into pENTR 5'-TOPO and assembled using a Gateway LR reaction with vector #302, vector #394 and the middle entry vector pME-MinPro-GFP to form the expression vectors DNaseIpeak-LCR-MinPro-GFP and Gata1peak-LCR-MinPro-GFP. To assemble the LCR without the DNase I peak, LCRP1/P2-MinPro-GFP, the fragment of the LCR 5' of the DNase I peak was amplified by PCR and cloned into pENTR 5'-TOPO (Invitrogen). The fragment of the LCR 3' of the

DNase I peak was amplified by PCR and cloned into pENTR 5'-TOPO (Invitrogen). The 5' fragment was then liberated by digest with BglII and BssHII (added to the primer sequences) and cloned, in the correct orientation, into the vector containing the 3' LCR fragment. This 5' entry vector was then assembled using a Gateway LR reaction with vector #302, vector #394 and pME-MinPro-GFP to form the expression vector LCRP1/P2-MinPro-GFP.

The middle entry vector pME-MinPro-GFP was constructed by PCR amplifying and inserting the minimal promoter from the mouse β-globin gene, previously used for enhancer screening in zebrafish (Woolfe et al., 2005), upstream of GFP (into the NcoI site of vector #383 [123]; NcoI sites added in primers: β-globin forward primer, GGCACCATGGCCAATCTGCTCAGAGAGGACA; β-globin reverse primer, GGCACCATGGGATGTCTGTTTCTGAGGTTGC).

All injections were carried out at the one cell stage. For the pEGFP-1 based vectors the linearized LCR-α/β_{a2}-GFP construct DNA or α/β_{a2}-GFP control DNA were injected at 100 ng/µl. Injections of the vectors assembled using the Gateway system were injected at 25 ng/ul in combination with Tol2 mRNA (15 ng/ul). The transiently injected embryos were observed at 24 hpf using a fluorescent microscope.

FACS Analysis of zebrafish peripheral blood

Zebrafish peripheral blood was isolated from deeply anesthetized adult α-LCR-α/β_{a2}-eGFP zebrafish as previously described [190], with the following modifications. Peripheral blood cells were placed into 200-300 uL of 0.9% PBS containing 5% fetal calf serum and 100 U/mL heparin, then filtered through a 40 µm nylon mesh to ensure a single cell suspension. Propidium iodide at 1 ug/mL (Sigma) was added as a marker to exclude dead cells and debris.

Fluorescence activated cell sorter (FACS) analysis was performed based on PI exclusion, forward scatter, side scatter, and GFP fluorescence using a FACS Vantage flow cytometer (Becton Dickinson, San Jose, CA). FACS data were analyzed using FloJo software.

DNase I hypersensitivity assay

Zebrafish peripheral blood was collected by cardiac puncture in anesthetized adult zebrafish. As the mature erythrocytes in lower vertebrates retain their nuclei, we are able to isolate erythrocyte nuclei from peripheral blood. Zebrafish liver tissues were dissected from anesthetized adult fish and cell suspensions were made using a glass tissue homogenizer in 1X PBS. DNase I hypersensitivity assays of both erythrocytes and liver cells were carried out as previously described (Sabo et al., 2006). Bowtie [191] was used to map DNase I sequence reads onto Zv9 and formatted into BAM files. MACS [192] was used to compute peaks and create browser tracks.

Briefly, the cells were treated with 0.025% Igepal in Buffer A (15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM Spermidine) for 8 minutes. The released nuclei were pelleted and incubated for 3 minutes at 37°C in DNase buffer (13.5mM Tris-Cl pH8.0, 88.5 mM NaCl, 54 mM KCl, 0.9 mM EDTA, 0.45 mM EGTA, 0.45 mM Spermidine, 6 mM CaCl2) plus 0, 40, 60, 80 or120 units/ml of DNase I (Sigma Aldrich). The reaction was stopped by the addition of an equivalent volume of Stop Buffer (50mM Tris-Cl pH 8.0, 100 mM NaCl, 0.10 %SDS, 100 mM EDTA) and incubated for 1 hr with 1100 units/ml Proteinase K. Small DNase I fragments from the 120 unit DNase I treatement were enriched using a sucrose gradient as previously described [193]. These ends were then labeled and amplified for Solexa sequencing as previously described [194]. Bowtie [191] was used to map DNase I sequence reads onto Zv9

and formatted into BAM files. MACS [192] was used to compute peaks and create browser tracks with the following input parameters: gsize = 15000000000; bandwidth = 50; and p-value = 1e-09.

ChIP-seq of zebrafish red cells

ChIP was performed [195], the sample sequenced and bioinformatic analysis performed [192] as previously described. Briefly, red cells from 10 adult zebrafish were isolated for each ChIP-seq reaction, cross-linked, prepared with the Illumina/Solexa Genomic DNA kit (Illumina-IP-102-1001), sequenced and analyzed using Model-based Analysis of ChIP-seq (MACS) [192].

Red cells from 10 adult zebrafish were isolated for each ChIP-seq reaction. The cells were isolated by puncturing the zebrafish heart with a heparinized tip and were transferred to a tube of 0.9% PBS, 5% fetal bovine serum, and 0.3U-1U heparin/ml. ChIP was performed as described in Lee et al.[195]. Cells were crosslinked with the addition of 1/10 of 11% formaldehyde diluted in Formaldehyde solution (50 mM HEPES-KOH, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 11% formaldehyde) for 5 min at room temperature. 1/20 of 2.5M Glycine was added to quench the formaldehyde and the cells were washed twice with ice cold PBS. Cells were either flash frozen in liquid nitrogen and kept at -80°C or processed immediately with the addition of 10 ml of lysis solution 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1× protease inhibitors) and incubation for 10min at 4°C. The cell lysate was centrifuged for 5min at 4°C at 1350g and the pellet was resuspended in 10ml of lysis solution 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mMEDTA, 0.5 mM EGTA, 1× protease inhibitors) and incubated for 10 min at room temperature. The cells were pelleted for 5min at 1350g at 4°C and then resuspended in 1ml

of Lysis buffer 3/Sonication buffer (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1× protease inhibitors). The lysate was sonicated in a Bioruptor sonicator for 25 cycles of 30sec with 59 sec intermission between each cycle. The lysate was centrifuged for 10 min at 18.000g at 4°C. The supernatant was transferred to a new tube and 10% Triton X was added. The lysate was added to 50 of Dynal beads that were crosslinked to GATA1 (Anaspec 55507) or H3K4me3 (Millipore 07473) antibody. 50 µl of lysate was kept for whole cell extracts. To crosslink the antibodies to the beads, 50 µl of beads were washed twice with PBS with 0.5% BSA and then 250µl of PBS/0.5%BSA together with 5µg of antibody was added to the beads. The beads were left to rotate for at least 6hrs to overnight at 4°C. After the incubation the beads were washed twice with PBS/BSA 0.5% and were ready to use.

After the addition of the cell lysate to the beads the tubes were left to rotate at 4°C overnight. Next day the beads were washed for 4 min each time in the following consecutive solutions. Beads are washed 1X with wash buffer 1 (20mM Tris-HCl pH8, 150 mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100), 1X with wash buffer 2 (20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1%Triton X-100), 1X with wash buffer 3 (10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40) and 1X with TE. All washes are performed at 4°C. The chromatin was eluted from the beads with the addition of 200µl elution buffer (50 mM Tris-Hcl, pH 8.0, 10 mM EDTA and 1% SDS) by heating at 65°C for 30 minutes in a shaking heat block. The eluted chromatin was transferred to a new tube and the crosslinking was reversed by heating at 65°C for 6hrs to overnight. The whole cell extracts were also reverse crosslinked by the same procedure after the addition of 150µl of elution buffer.

After the reversal of crosslinking RNAse was added to the lysate at a final concentration of 0.2mg/ml for 2hrs at 37°C, followed by the addition of proteinase K at a final concentration of 0.2µg/ml for 2hrs at 55°C. The DNA was recovered by Phenol/Chloroform/Isamylic alcohol precipitation and resuspended in 30µl of H2O for the ChIP and 150µl for the whole cell extracts.

ChIP-seq sample preparation [196]

The samples were prepared with the Illumina/Solexa Genomic DNA kit (Illumina- IP-102-1001) according to the manufacturer's instructions. Briefly 200ng of Input DNA and 50ng of ChIP DNA are used. The first step was to turn the DNA overhangs into phosphorylated blunt ends. The sample was then purified with a Qiagen PCR purification kit. The second step of the library preparation was the addition of a single A in the 3' end to allow for directional ligation. Samples were purified again with the Qiagen MimElute kit. Next, Illumina adapter oligonucleotides (1/100 dilution) were added to the samples, followed by purification of the samples with the Qiagen PCR purification kit. The samples were amplified by PCR (limited amplification to 18 cycles) that added additional linker sequence to the fragments to prepare them for annealing to binding sites in the Genome Analyzer flow-cell. The amplified samples were then separated on a 2% agarose gel and products between 150-350 bp were selected and extracted from the gel (the products include fragments of 50-250 bp plus approximately 100 bp of primer sequence). All protocols for Illumina/Solexa sequence preparation, sequencing and quality control are provided by Illumina (http://www.illumina.com/pages.ilmn?ID=203).

Polony generation and sequencing

The DNA library (2-4 pM) was applied to the flow-cell (8 samples per flow-cell) using the Cluster Station device from Illumina. The concentration of library applied to the flow-cell

was calibrated such that polonies generated in the bridge amplification step originate from single strands of DNA. Multiple rounds of amplification reagents were flowed across the cell in the bridge amplification step to generate polonies of approximately 1,000 strands in 1 μm diameter spots. Double stranded polonies were visually checked for density and morphology by staining with a 1:5000 dilution of SYBR Green I (Invitrogen) and visualizing with a microscope under fluorescent illumination. Validated flow-cells were stored at 4°C until sequencing. Flow-cells were removed from storage and subjected to linearization and annealing of sequencing primer on the Cluster Station. Primed flow-cells were loaded into the Illumina Genome Analyzer 1G. After the first base was incorporated in the Sequencing by-Synthesis reaction the process was paused for a quality control checkpoint. A small section of each lane was imaged and the average intensity value for all four bases was compared to minimum thresholds. Flow-cells with low first base intensities were reprimed and if signal was not recovered the flow-cell was aborted. Flow-cells with signal intensities meeting the minimum thresholds were resumed and sequenced for 26 cycles.

Bioinformatics analysis of ChIP-seq data

Samples were sequenced and mapped to ZV9 at the Whitehead Institute Genome Technology Core. Peaks were detected for nominally 36 bp short reads passing Eland quality control using Model-based Analysis of ChIP-seq (MACS) [192] with a bandwidth of 50, mfold of 15, and p-value of 1e-09. The resulting WIG files from MACS were used for visualizing the results.

In situ hybridization

The whole-mount *in situ* hybridization protocol was carried out as previously described [197] using antisense probes amplified from digested plasmids.

Accession numbers

The Dnase I and ChIP-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [198] and are accessible through GEO Series accession number GSE35895 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35895).

Results

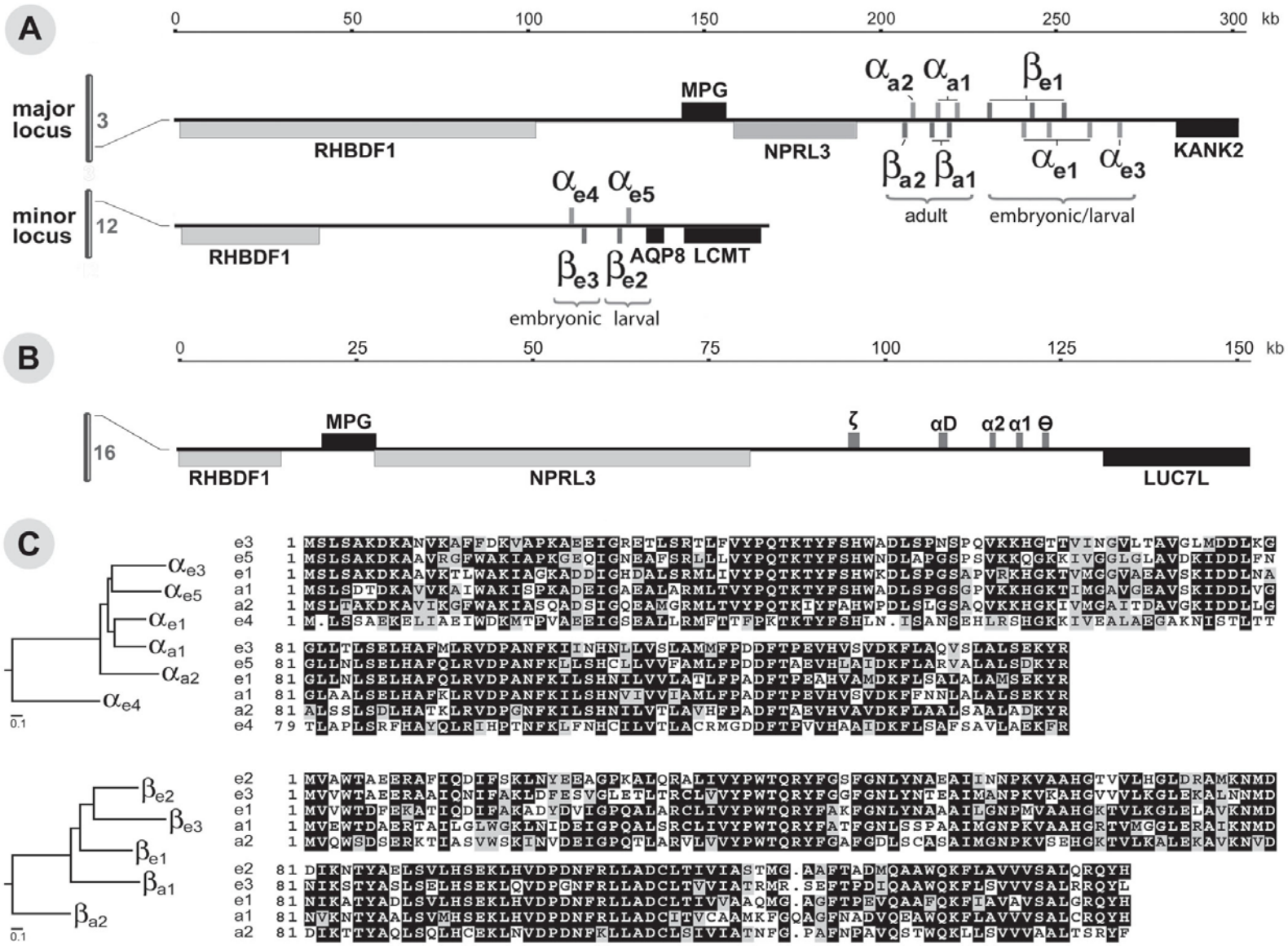
Analysis of the genomic regions of the globin gene loci

The genes encoding the zebrafish globins reside on two separate chromosomes. Chromosome 3 contains the major *globin* locus with 13 globin genes and chromosome 12 houses the minor globin locus with four *globin* genes (Figure 2-1A). The sequence from the major globin locus was assembled from fully sequenced BACs and phage artificial chromosomes (PACs) available from the zebrafish genome sequencing project at the Sanger Genome Sequencing Center (GenBank ID: AC103581, GenBank ID: AL953863, GenBank ID: BX004811 and GenBank ID: CU464181) based on overlapping sequences among these genomic clones and their encoded globins. Additional sequenced BACs contain different numbers of *globin* genes, such as AL845551 and AL929176, indicating the presence of various globin locus haplotypes within the population. Figure 2-1A likely represents a summary of more than one haplotype of the globin region in individual zebrafish that were used to make BAC and PAC libraries (BUSM1, CHORI, CH73, and DanioKey). The major globin locus is syntenic to human

Figure 2-1. Major and minor zebrafish chromosomal loci and protein sequences.

(A) The major and minor zebrafish *globin* loci are located on chromosome 3 and 12 respectively. The major locus was assembled by aligning bacterial artificial chromosomes (BAC) and phage artificial chromosomes (PAC), AC103581, AL953863, BX004811 and CU464181, and contains 13 *globin* genes. The minor locus was assembled by aligning 2 BACs, CR352324 and BX572076, and contains 4 *globin* genes. The timing of expression is denoted by "embryonic," "embryonic/larval," and "adult."The surrounding regions included are syntenic with other teleost and mammalian species ("Results"). The scale indicated the distance of the region from the beginning of *rhbdf1* on both loci. (B) The human *α-globin* locus, adapted from Higgs and Wood [182], is syntenic with both the major and minor zebrafish *globin* loci. (C) Analysis of the similarity of the various *globins*, broken into *α-* and *β-globins*, by protein sequence. Phylogenic analysis was performed using the Clustal W method (MegAlign; DNASTAR). Shades of gray indicate level of conservation at the amino acid level.

Figure 2-1. (Continued)



chromosome 16 which contains the *α-globin* locus (Figure 2-1B). The adjacent genes around the zebrafish major locus share many similarities with the human α-globin locus [33,34]. These include *rhbdf1* (*c16orf8*), *mpg*, *nprl3* (*c16orf35*), and the *kank2* (*flj20004*; an ankyrin-like gene). The LCR is found within an intron of the adjacent nprl3 gene [199]. There are several differences in the zebrafish major locus compared to mammals. First, genes are primarily found in *α/β-globin* pairs with a head-to-head orientation. This gene distribution implicates a primitive mechanism to coordinate equal expression of the two globin genes: utilization of a single common promoter and shared enhancer. Second, in the zebrafish the adult globin cluster is located closer to the putative locus control region ("The Putative α-LCR and Proximal *globin* Promoter Confers Erythroid Specific Expression") than the embryonic/larval cluster. This is in contrast to humans and mice where the globin genes are expressed in the order of their position within the locus, with the earliest expressed globins closest to the LCR [38].

The minor locus is covered on two BACs, CR352324 and BX572076, and each contains the same four globins. This locus is less syntenic with the mammalian globin loci, while remaining syntenic with other fish globin loci. Similar to *Fugu* it shares the *rhbdf1* gene that is also present at the major globin locus [34]. This locus is flanked on the other side by the *aqp8* and *lcmt* genes (Figure 2-1), with the *lcmt* gene also being a conserved flanking gene with *Fugu* [34]. Like the major locus, genes are also arranged in pairs; however, only the intermediate expressed globins α_{e5} and β_{e2} are arranged in a head-to-head orientation. The α_{e4} and β_{e3} pair is arranged in a tail-to-tail fashion.

Genome wide survey of globin coding loci, on Zv9 and available BAC sequences, has identified 17 coding regions that harbor conserved *globin* genes (Table 2-2). Through sequence similarity analysis at both the nucleotide and amino acid levels (Figure 2-1C), 11 unique globin

Table 2-2. Unique identifiers and chromosomal coordinates for zebrafish *globin* genes.

Gene	NCBI Official Symbol	Sequence Accession #	ZFIN ID	Chromosome Coordinates	Ensemble Gene ID
$\alpha_{a1a/b}$	*hbaa1*	NM_131257 NM_001033093	ZDB-GENE-980526-79	3:55951784-55952605 3:55956810-55957634	ENSDARG00000079078
α_{a2}	si:ch211-5k11.6	NM_001013461	ZDB-GENE-081104-38	3:55945257-55946071	ENSDARG00000069735
$\alpha_{e1a/b/c}$	*hbae1*	NM_182940	ZDB-GENE-980526-80	3:55972039-55972738 3:55979221-55979920 3:55990381-55990796	ENSDARG00000089475
α_{e3}	*hbae3*	BC071507	ZDB-GENE-990706-3	3:55998262-55999416	ENSDARG00000079305
α_{e4}	zgc:163057 (*hbq1*)	NM_001082834.1	ZDB-GENE-070410-131	12:21688580-21689767	ENSDARG00000045144
α_{e5}	LOC561790 (*hbz*)	EH450566	ZDB-GENE-101028-1	12:21704893-21706037	ENSDARG00000045142
$\beta_{a1a/b}$	*ba1*	NM_131020	ZDB-GENE-990415-18	3:55955117-55955918 3:55950129-55950930	ENSDARG00000089087
β_{a2}	zgc:92880	NM_001003431	ZFIN:ZDB-GENE-040801-164	3:55942902-55943621	ENSDARG00000069734
$\beta_{e1a/b/c}$	*hbbe1.1*	NM_198073	ZDB-GENE-030616-7	3:55974334-55975111 3:55965769-55966546 3:55982961-55983724	ENSDARG00000089963
β_{e2}	*hbbe2*	NM_212846	ZDB-GENE-040702-1	12:21702360-21703066	ENSDARG00000045143
β_{e3}	*hbbe3*	NM_001015058	ZDB-GENE-980526-287	12:21691962-21693036	ENSDARG00000038147

genes were identified with evidence from existing cDNA sequences. Several of the genes have been duplicated or triplicated and share identical sequence at the nucleotide and amino acid level. Comparative analysis of their proximal promoters also shows that they are identical (data not shown). The majority of these gene species have been previously identified in other studies for embryonic [30] and adult zebrafish globins [32]. Based on phylogenetic analysis, these globins are grouped into *α-* and *β-globin* branches. Based on sequence diversity of BACs containing globin genes, comparison of different strains reveals the presence of different haplotypes of adult globins (data not shown). New globins have been identified on the minor locus using a sequence search. These *globin* genes, α_{e4}, α_{e5}, β_{e2} and β_{e3}, were not previously analyzed for their sequence similarity and gene expression. Both β_{e2} and β_{e3} group with the previously known zebrafish *β-globin* genes based on amino acid sequence. Similarly, α_{e5} groups with the previously known zebrafish *α-globin* genes. α_{e4} is the most divergent of all the newly analyzed globin genes and, apart from teleosts, is the most similar to the α^D globin found in chickens. The unique number of globins present and similar organization of the zebrafish *globin* loci is mostly likely the result of genome duplication(s).

Globin gene expression during development identifies two *globin* switches

To characterize the developmental *globin* switching process, the expression of the *globin* genes was followed in embryos, juveniles and adult fish up to 32 days post fertilization (dpf), and then again assessed at one year of age. Total RNA samples were analyzed with quantitative real-time PCR (qPCR) primers designed to amplify specific *globin* species.

The *globin* expression studies delineate three stages of expression (Figure 2-2). The embryonic stage is defined by the expression of β_{e3}, the only exclusively embryonic *globin*, with

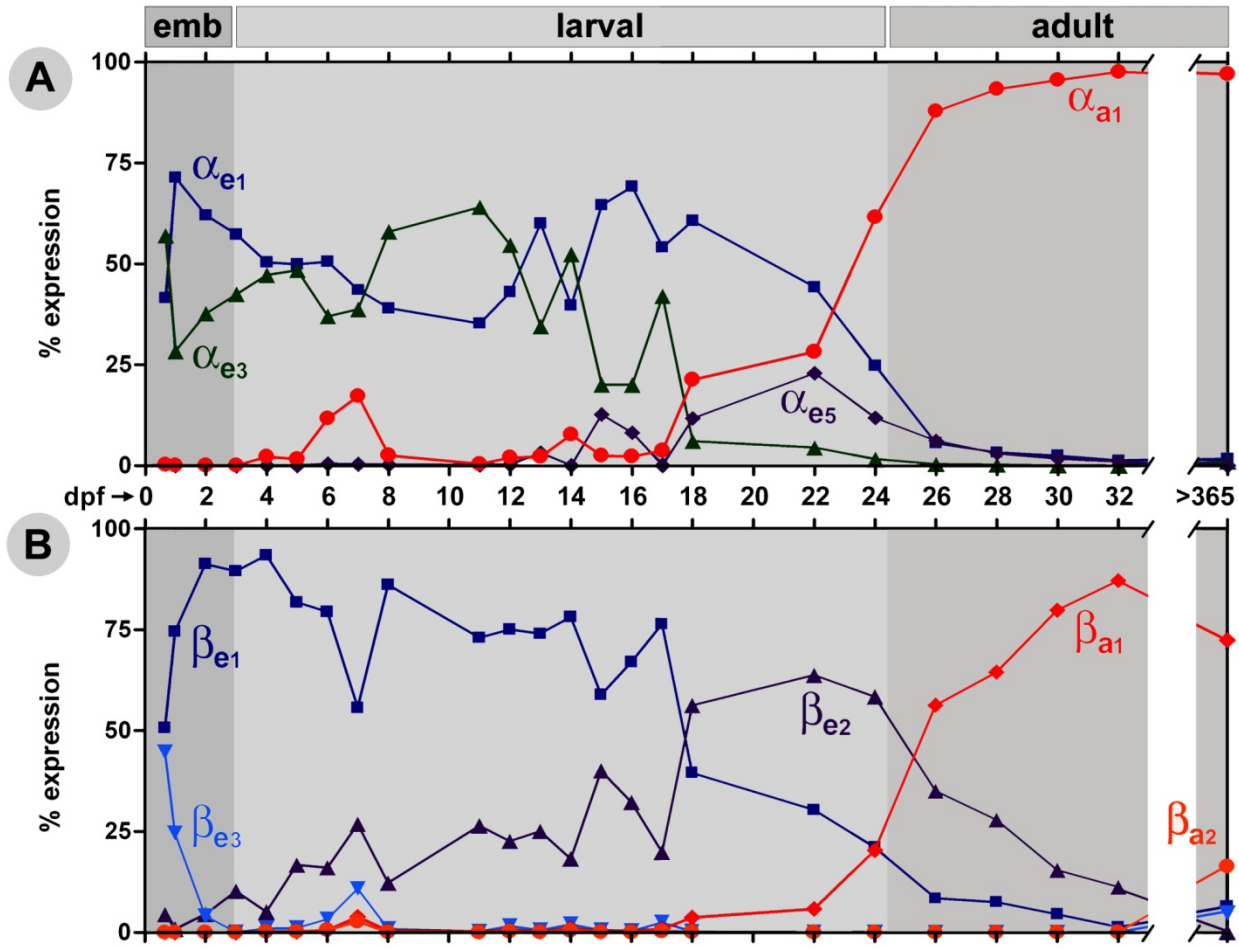


Figure 2-2. Relative *globin* expression levels throughout development.

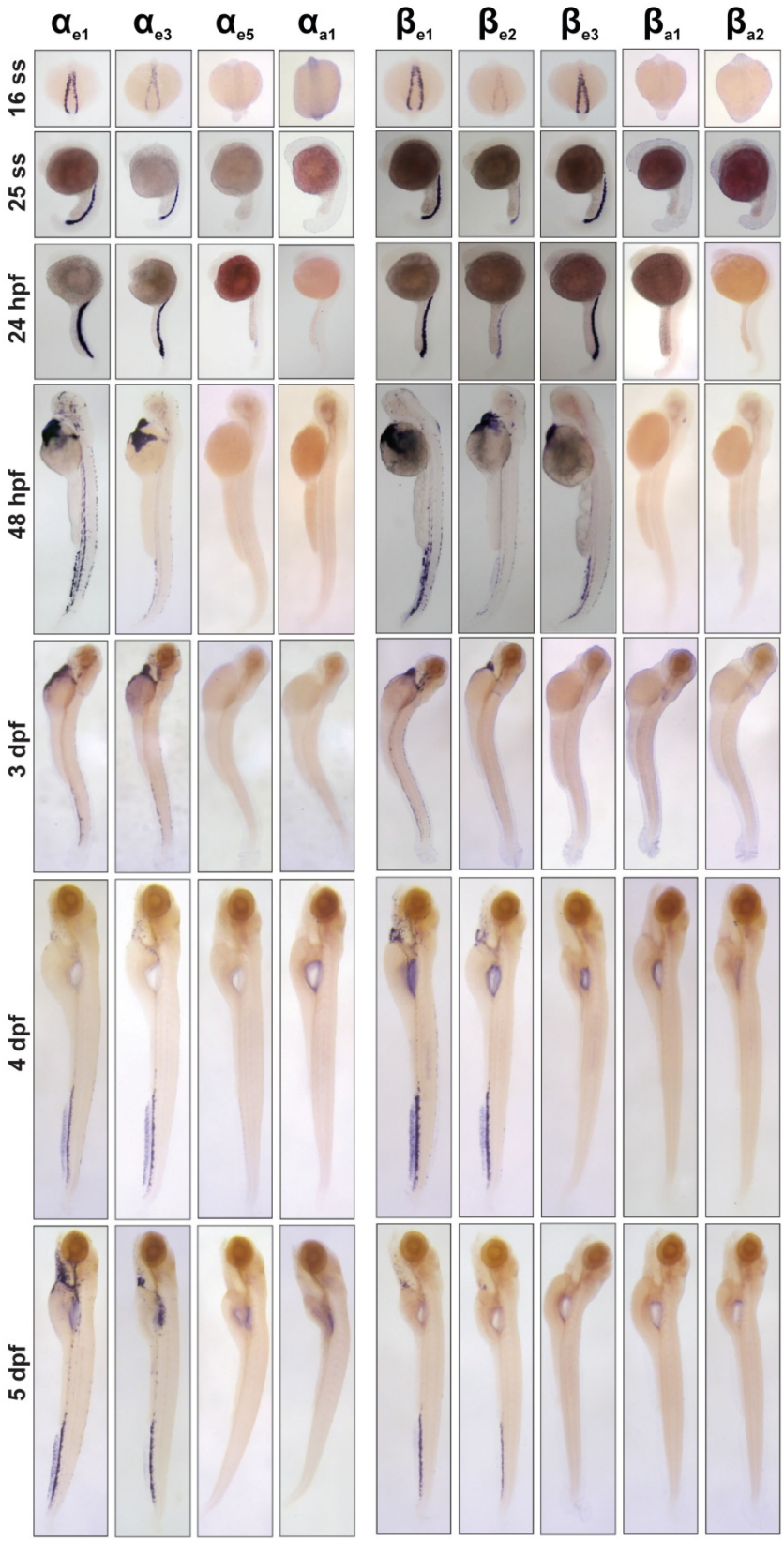
The relative expression level changes of the *α-globin* (A) genes and the *β-globin* genes (B) in the zebrafish are shown. In both A and B, the approximate embryonic (emb), larval and adult stages of *globin* expression are denoted by shades of gray. The first time point depicted is the 16 somite stage. Relative expression levels were determined by quantitative real-time PCR and normalized to *band3*. The *globin* genes α_{e4} and α_{a2} are not depicted as no significant expression was detected at any of these time points.

substantial contribution from α_{e1}, α_{e3} and β_{e1} (Figure 2-2 "emb" panel). The first switch then appears to begin between 24 hpf and 36 hpf, marked by the sharp decrease in β_{e3}. The larval stage is characterized by increasing expression of β_{e2} and, in the later portion, α_{e5}, two genes oriented head-to-head in the genome in the minor locus (Figure 2-1), as well as the maintained expression of many of the *globins* from the embryonic stage. These embryonic *globins* begin to decrease as the larval *globins* peak near the end of this period. The second switch that establishes the mature, adult *globin* expression, as defined by the expression pattern observed at 1 year of age, is characterized by the decreasing expression of the embryonic and larval *globins* and the increasing expression of the nearly exclusively adult *globins* α_{a1} and β_{a1} (Figure 2-2). This switch begins around 22 dpf, with the continued decline of the embryonic/larval stage *globins* and the start of the decline of the larval *globins* α_{e5} and β_{e2}. The adult *globin* expression pattern is nearly completely established by 32 dpf. The genes from the cluster closest to the Locus Control Region (LCR), α_{a1} and β_{a1}, contributing the majority of hemoglobin for the adult fish and a smaller contribution from β_{a2}. For the first 5 dpf, we examined the expression of the *globin* genes by *in situ* hybridization (Figure 2-3). These results support the qPCR results and provide further evidence there is a *globin* switch from the embryonic to larval stages in the zebrafish. The decrease of *βe3* expression between 48 hpf and 3 dpf observed in the qPCR data (Figure 2-2) is also observed by *in situ* hybridization as a loss of the staining between these two time points (Figure 2-3). The increase in the expression of *βe2* between the 16 s.s. and 5 dpf identified by qPCR is also evident in the increased intensity of the staining over the course of these time points by *in situ* hybridization. Together, these gene expression studies demonstrate that the zebrafish has an embryonic to larval switch and a larval to adult *globin* switch.

Figure 2-3. G*lobin* expression by *in situ* hybridization through 5 dpf.

Expression patterns of the *α*- (A) and *β-globin* (B) genes. For detected genes, expression is seen in bilateral stripes at the 16ss, in the ICM at the 25ss and 24 hpf stage. With the onset of circulation at 24 hpf, *globin* positive cells can be seen throughout the vasculature with probes for expressed *globin* genes, particularly in the vascular plexus of the caudal hematopoietic tissue.

Figure 2-3. (Continued)



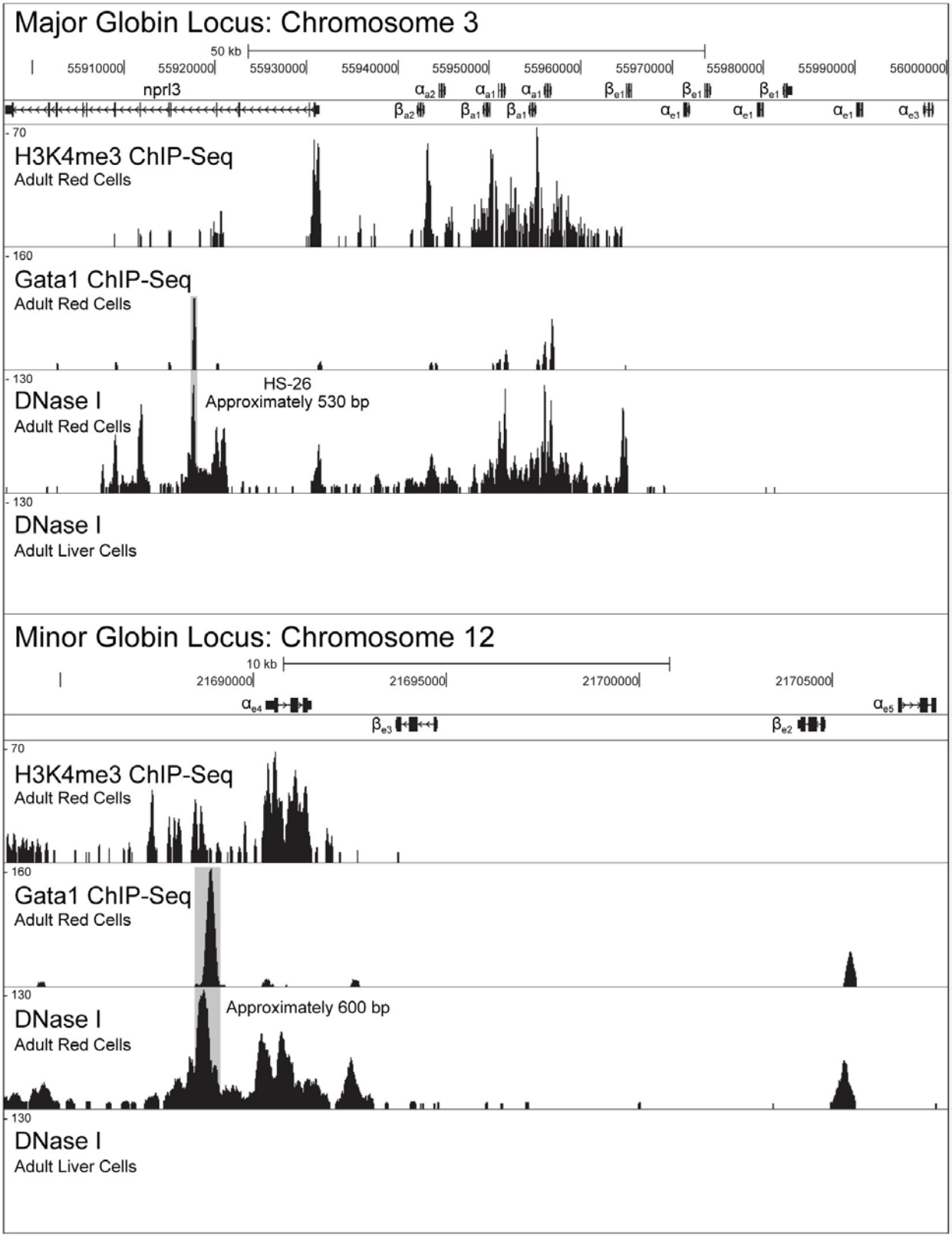
Analysis of the genomic structure of the *globin* loci

To define long-range *globin* enhancers and promoters, DNase I hypersensivity mapping, the active H3K4 trimethylation (H3K4me3) mark was mapped and ChIP-seq analysis for the canonical erythrocyte transcription factor GATA1 [40] was carried out in mature adult erythrocytes, which are nucleated in the zebrafish. DNase I and H3K4me3 peaks were observed at expected genes such as GATA1, SCL and KLF4 in comparison to the control liver cells (Figure 2-4). As expected, the GATA1 ChIP-seq showed distinct peaks corresponding to the proximal promoters of the transcriptionally active adult *globin* genes. Within both the major and minor *globin* loci, the HS are highly correlated with H3K4me3 within gene bodies, and the GATA1 sites show a restricted binding correlated with strong HS signals (Figure 2-4). HS and H3k4me3 signals are evident through the body of the adult *globins* in the major locus and generally absent from those of the embryonic *globins*, consistent with the, respectively, high and low levels of expression of these gene in adult red cells (Figure 2-2). These data indicate an open genomic structure at the adult *globin* promoters. Within the *globin* gene cluster, the binding of GATA1 appears to be localized to the proximal promoters of the highly expressed adult *globin* genes in conjunction with elevated HS signal and H3K4me3 signal. An additional strong HS peak that does not correspond with a GATA1 signal appears near the proximal promoter for β_{e1}, the embryonic gene closest to the transcriptionally active adult *globins*. As observed in other organisms, this HS may be playing a role in repressing the subsequent *globins* [200]. A similar pattern is observed in the minor locus (Figure 2-4); the only coding region with high levels of H3K4me3 and HS is α_{e4}, for which no expression was detected. A peak within the repressed region, as determined by qPCR expression (Figure 2-2), of β_{e2} and α_{e5} is also observed.

Figure 2-4. Analysis of the chromosomal state of the major and minor *globin* loci.

All tracks were mapped to Zv9 on the UCSU genome browser (http://genome.ucsc.edu/). In some cases annotated genes were renamed in order to adhere to the naming convention, and in cases where a *globin* gene was not annotated, the UCSC BLAT tool was used to locate the ORFs included in the figure. The shaded gray areas indicate the confirmed and putative regulatory regions in the major and minor loci respectively.

Figure 2-4. (Continued)



These results are specific to red blood cells as the control liver cells do not exhibit any of these patterns.

In addition to the HS, H3K4me3 and GATA1 sites within the *globin* gene cluster itself, a number of putative upstream regulatory elements in both loci were identified. The gene adjacent to the *globin* coding sequences of the major locus, *nprl3*, which is also expressed in red cells, contains a number of HS. One of these peaks spans a region of approximately 530 bp and contains a core region of approximately 200 bp that features NFE2, CACCC and functional GATA1 motif binding sites. This region is located approximately 26 kb upstream of the nearest *globin* coding sequence (HS-26; gray box, Figure 2-4 "Major Globin Locus"). This region shares similarity to MCS-R2 (HS-26 in mice and HS-40 in humans) [45]. The overall genomic structure of the minor locus appears to be less homologous to *globin* loci in higher vertebrates, but a number of putative regulatory sites were able to be detected. Three peaks, two with strong H3K4me3 and HS signal and one HS without H3K4me3 signal, are spread across the locus (Figure 2-4). The first is located upstream of α_{e4}, the second spanning the coding sequence of the gene and the third between β_{e2} and α_{e5}. All three contain numerous GATA1-, NF-E2- and CACC- binding motifs. As in the major locus, the GATA1 binding is mainly localized to proximal promoters, though the genes are not active (Figure 2-4). The strongest DNase I hypersensitive peak in the minor locus (gray box, Figure 2-4 "Minor Globin Locus"), located upstream of all the coding sequences, contains the characteristic binding motifs and coordinates with the strongest GATA1 peak, and it is therefore most likely the LCR for the minor locus. These measures of the genomic structure of the *globin* locus strongly indicate a role for these features in the overall regulation of *globin* expression and demarcating putative regulatory elements.

The Putative α-LCR and Proximal *globin* Promoter Confers Erythroid Specific Expression

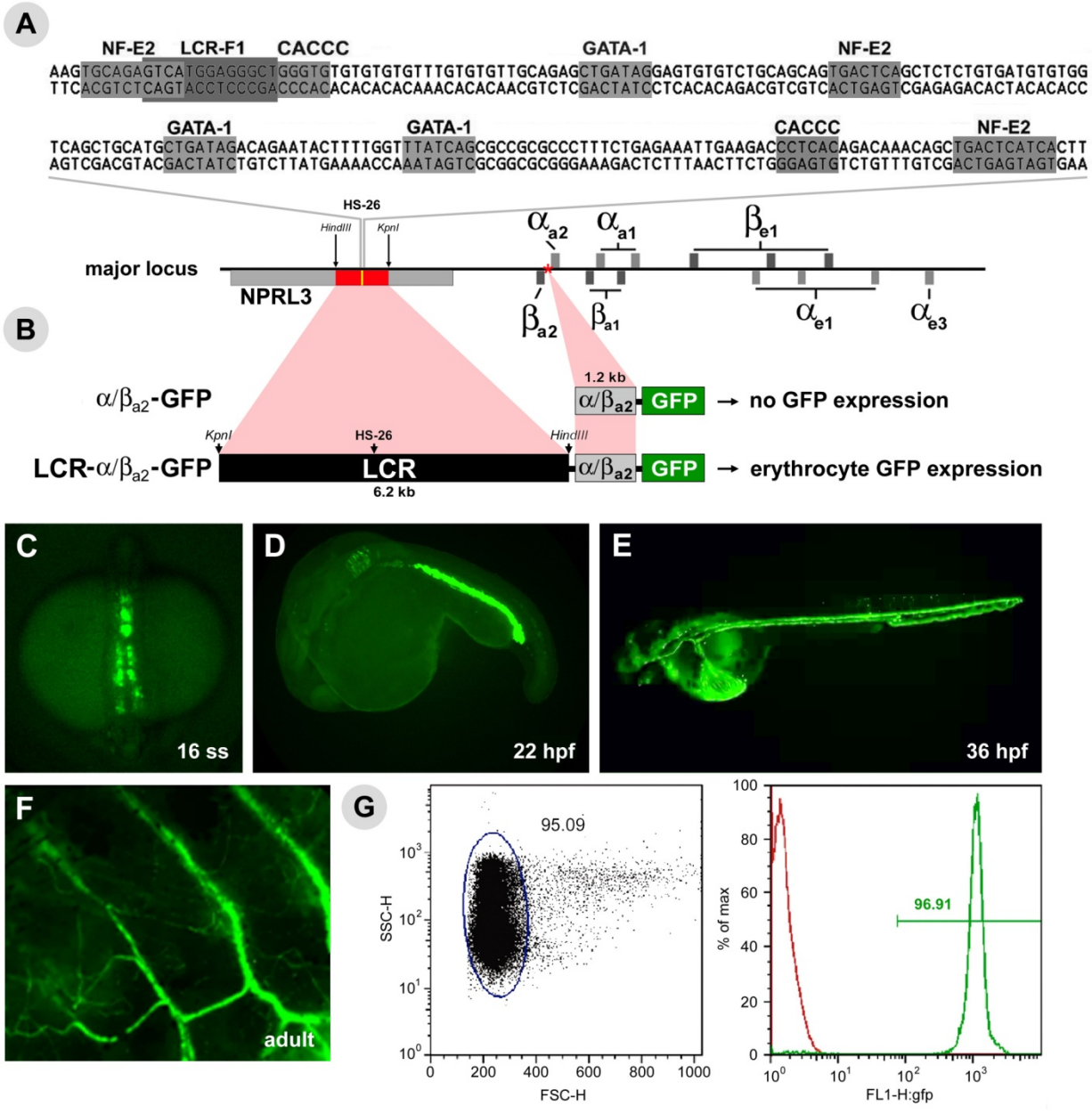
Due to the open chromatin and high degree of sequence homology to known *globin* regulatory regions (MCS-R2) the role of the HS-26 peak in the regulation of *globin* gene expression in zebrafish was further investigated. The MCS-R2 regulatory site has been mapped in humans to a 300 bp region within the 5th intron of the *nprl3*, that contains GATA(1)-, NF-E2/AP1- and CACC-binding motifs [38,182]; both the synteny and binding motifs are conserved in the zebrafish homologue, HS-26 (Figure 2-5A). A similar region was identified through comparison to additional species [45,201]. This region has been identified as being sufficient for robust and specific reporter expression in red blood cells [40].

In order to functionally test the ability of the putative zebrafish regulatory region HS-26 to confer robust, erythroid-specific expression *in vivo*, a fragment from a bacterial artificial chromosome (CH211-113F11; AL953863) containing HS-26 from the 5th intron of *nprl3* as well as a portion of the α/β_{a2} bidirectional proximal promoter were cloned into GFP reporter constructs and assessed their expression *in vivo* (Figure 2-5B). Transgenic embryos were grown to adulthood and transgenic lines established. Multiple lines were monitored and found to have robust erythroid specific visible GFP expression beginning at the 16 somite stage in bilateral stripes in the posterior mesoderm (Figure 2-5C). At 22 hpf, the ICM region, the site of primitive hematopoiesis, is label with GFP (Figure 2-5D). At 24 hpf, these GFP positive cells enter circulation and persist through adulthood (Figure 2-5E, F). The specificity of the GFP expression to mature erythrocytes was confirmed by FACS analysis of adult zebrafish peripheral blood (Figure 2-5G). GFP expression in erythrocytes is visible from the 16 somite stage to adulthood, demonstrating that the reporter construct does not "switch." In addition, expression of the GFP mRNA was assessed in order to directly compare the expression of the endogenous

Figure 2-5. Generation and characterization of LCR-GFP transgenic zebrafish line.

(A) Sequence and location of the HS-26 DNase I hypersensitivity site (HS-26; yellow rectangle) and locus control region (LCR; red rectangle) within the major *globin* locus. The proximal *globin* promoter used is indicated with a red star. Shaded boxes indicate binding motifs for canonical erythrocyte transcription factors. (B) Representation of the vectors assembled from genomic fragments and the resulting GFP expression patterns. Fluorescent images of LCR-GFP transgenic zebrafish embryos at 16ss, 22 hpf and 36 hpf respectively (C-E). (F) Fluorescent image of the flank of an adult LCR-GFP transgenic zebrafish. (G) Red blood cell gate as determined by forward and side scatter for peripheral blood, and the analysis of the percent of GFP positive red blood cells in an LCR-GFP transgenic adult (green) versus in a wild-type adult (red).

Figure 2-5. (Continued)



globins versus that of the transgene. GFP mRNA expression can be detected beginning at approximately the 12 somite stage, which coincides with the onset of endogenous *globin* expression [30] (Figure 2-6).

Additional constructs were assembled to confirm HS-26 as the functional *globin* enhancer. The various constructs were injected at the one cell stage, the embryos were allowed to develop until the 24 hpf stage and observed under a fluorescent microscope. The ability of the α/β_{a2}-bidirectional promoter to drive erythroid expression in the absence of the enhancer was tested by injecting the reporter construct without the putative LCR region. None (0%; 0/50) of the injected embryos expressed visible GFP expression, while 92% (22/25) of those injected with the construct containing the full LCR expressed GFP (data not shown). Two constructs were designed with a truncated LCR region containing either HS-26, as defined by the bounds of the DNase I peak (Figure 2-4), or the fraction of HS-26 located under the GATA1 peak (Figure 2-4) driving GFP under the control of the α/β_{a2}-bidirectional promoter. Injection of each construct, in parallel with an mCherry reporter plasmid to control for injection efficiency, showed robust ICM GFP expression in 100% (50/50; 50/50) of productively injected embryos. Replacing the zebrafish α/β_{a2}-bidirectional promoter with a minimal promoter derived from the mouse *β-globin* locus confirmed the enhancer ability of both the DNase and GATA1 peaks. GFP expression was observed in the ICM of 100% (50/50) of embryos productively injected with the GATA1 peak enhancer/minimal promoter construct and 94% (47/50) of embryos injected with the DNase peak/minimal promoter construct (Figure 2-7). Conversely, 82% (41/50) of embryos productively injected with LCRP1/P2-MinPro-GFP, which contains the full LCR except for the DNase I peak, did not express GFP in the ICM. This demonstrates that the functional component of the full LCR region is contained within the DNase I peak. The ability of HS-26 to confer

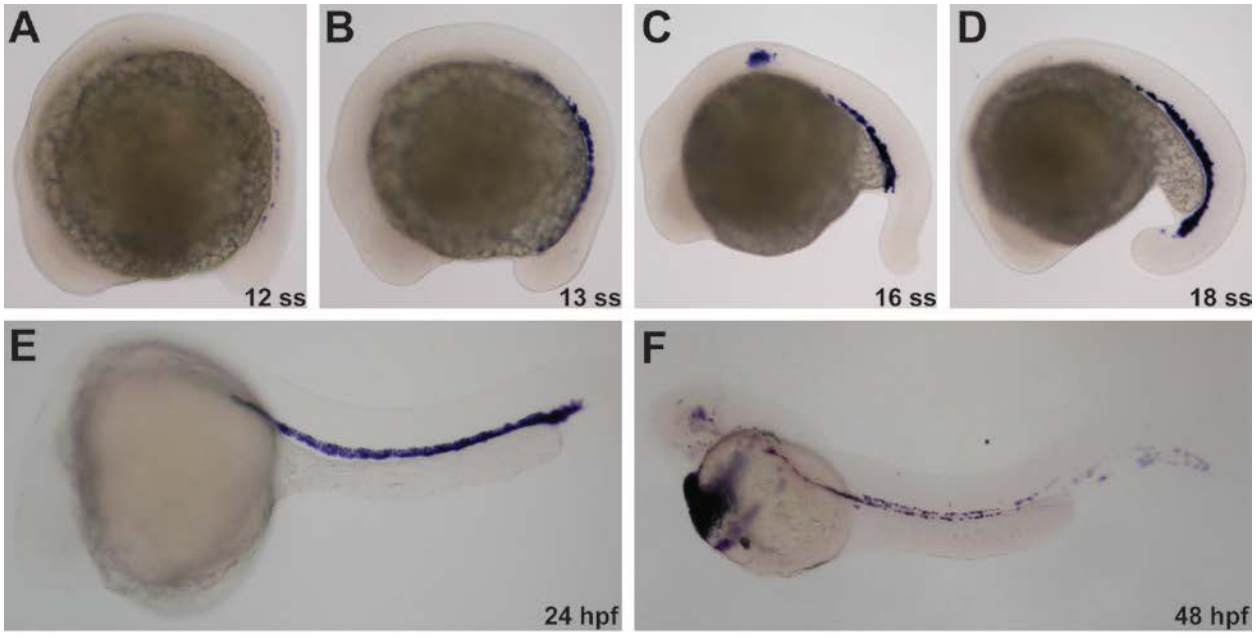


Figure 2-6. GFP mRNA expression in LCR-GFP transgenic zebrafish line.

In situ hybridization was performed to detect GFP RNA expression in the LCR-GFP transgenic zebrafish line as RNA expression precedes visible GFP fluorescence. This was performed at the 12 s.s. (A), 13 s.s. (B), 16 s.s. (C), 18 s.s. (D), 24 hpf (E) and 48 hpf (F) stages.

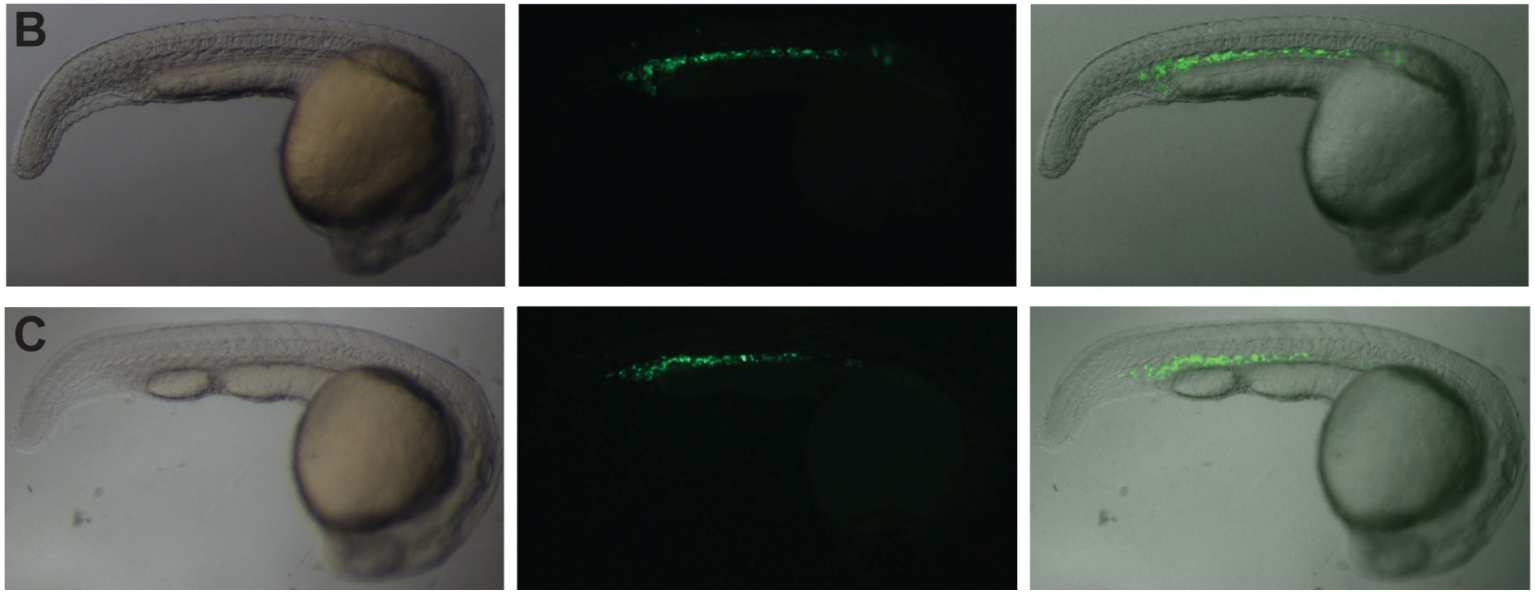
Figure 2-7. Functional characterization of the LCR in transiently injected embryos.

(A) Summary of the transgenic constructs injected including the enhancer element, promoter and vector backbone. All vectors listed contain a GFP reporter open reading frame. Bright field, GFP and composite images of representative Gata1Peak-LCR-MinPro-GFP (B) and DNaseIPeak-LCR-MinPro-GFP (C) embryos. [1]The # Observed refers to all embryos observed from the injected clutch. [2]The # Observed refers to the number of productively injected embryos as determined by expression by an mCherry reporter plasmid.

Figure 2-7. (Continued)

A

Vector Name	Enhancer	Promoter	Backbone	# GFP+/# Observed
LCR-α/β_{a2}-GFP	LCR	α/β_{a2}	pEGFP-1	22/25[1]
α/β_{a2}-GFP	None	α/β_{a2}	pEGFP-1	0/50[1]
Gata1Peak-LCR-α/β_{a2}-GFP	Gata1 Peak	α/β_{a2}	394	50/50[2]
Gata1Peak-LCR-MinPro-GFP	Gata1 Peak	Mouse Minimal	394	50/50[2]
DNaseI Peak-LCR-α/β_{a2}-GFP	DNaseI Peak	α/β_{a2}	394	50/50[2]
DNaseI Peak-LCR-MinPro-GFP	DNaseI Peak	Mouse Minimal	394	47/50[1]
LCRP1/P2-MinPro-GFP	LCRP1/P2	Mouse Minimal	394	9/50[2]



robust, specific expression *in vivo* confirms that the region is the LCR for the major *globin* locus and its functional homology to MSC-R2.

Discussion

Globin gene regulation serves as a paradigm for the study of gene expression. The zebrafish offers a genetic model to study globin gene switching in vivo. Although the erythroid program in zebrafish utilizes many of the same transcription factors as mammals, the unique structure of the fish globin loci ensures globin chain balance by regulating pairs of globin genes. Our analysis of the *globin* locus in zebrafish demonstrates a high level of synteny between the teleost *globin* loci and the mammalian α-*globin* locus. The LCR is very well conserved based on functional studies. The switching patterns include a switch from embryonic to larval, and from larval to adult globins. Surprisingly, the genomic structure of the loci and the binding of a canonical erythrocyte transcription factor are very similar between mammals and fish, despite the overall lack of primary DNA sequence conservation. This suggests that the overall genomic structure, including the binding of transcription factors, is more critical for globin gene expression than strict sequence conservation.

Our work supports the hypothesis that the ancestral locus is an *α-globin* locus [33,34]. The presence of both α- and β-like *globin* genes on both the major and minor loci may place the zebrafish closer to this ancestral locus than the puffer fish (*Fugu rubripes*), where one locus contains only α-like genes, more similar to the organizational structure found in mammals. The arrangement of the *globin* genes within the loci also implies a bi-directional promoter located between an α- and β-like gene as an ancestral mechanism for obtaining comparable levels of α- and β-like Globin protein. This is also supported by the localization of the GATA1 peaks (Figure 2-4) to these putative promoter regions in major locus. The mammalian *globin* loci are

arranged such that, for the predominant *globins*, the genes are temporally expressed in the order in which they are present in the genome [40,45]. Within the zebrafish genome, the genes are grouped into clusters of embryonic, embryonic/larval, larval and adult gene expression, and while the importance of physical location with respect to the LCR is conserved, the orientation is reversed for the major locus (Figures 2-1). Globin switching occurs in fish and utilizes the LCR elements to interact with specific regulatory elements near the individual globin genes.

The conservation of the regulatory network responsible for controlling *globin* switching, including the conservation of the transcription factors, the primary sequence of their binding site, the location of these site and their affects on transcriptional effect on other regulators and the *globin* gene themselves, establishes the similarity between the overall process of *globin* production in zebrafish and higher mammals. At approximately 16 s.s. in the bilateral stripes of the developing embryo, the primitive erythrocytes have a high expression of the embryonic *globin* genes α_{e1}, α_{e3}, β_{e1} and β_{e3} (Figure 2-2, 2-3). Around 24 hours post fertilization (hpf), primitive proerythroblasts enter circulation [27], expressing the same embryonic *globin*s, but with ratios different than those present at earlier time points. These cells presumably undergo "maturational" *globin* switching [26], as the *globin* expression of the embryo continues to change prior to the emergence of red blood cells derived from the next wave. These cells continue to mature in circulation and are the only circulating red cell population through 4 dpf [28], but can contribute past 7 dpf [27]. Between 1 dpf and 2 dpf, the ratios of the embryonic *globin*s previously expressed continue to change, and through 4 dpf β_{e3}*'s* contribution drops while β_{e2} expression increases. These changes coincide with the emergence of the erythromyeloid progenitor (EMP) population, which is responsible for generating the next wave of red blood cells to enter circulation [202]. Mature cells begin to enter circulation about 36 hpf, but

contribute to the globin expression of the whole embryo prior to this. Through this time period maturational switching as well as the evolving ratio of primitive to EMP cells are most likely affecting the global globin profile. The definitive hematopoietic stem cells are specified as early as 24 hpf but do not contribute to the mature red blood cell pool until approximately 10-14 dpf, having already migrated from the AGM to the caudal hematopoietic tissue (CHT), but may be contributing to the whole embryo globin profile prior to entering circulation. HSCs arriving from the CHT begin to seed the pronephros beginning around 4 dpf, with the definitive hematopoietic activity detectable short thereafter. This will provide full multilineage hematopoietic support for the remainder of the animal's life [203]. Therefore, it is likely that the precipitous decrease in the embryonic *globins* between approximately 17 dpf and 26 dpf (Figure 2-2) is the result of a decreasing contribution from the EMP wave erythrocytes, and an unknown contribution from maturational switching. The larval *globin*s α_{e5} and β_{e2} peak at a distinct time period. Definitive erythrocytes may not undergo maturational *globin* switching as the increase in α_{a1} and β_{a1} coincides approximately with the increasing contribution of definitive erythrocytes to circulation. The overlapping of these distinct cell populations in the embryo as well as remaining gaps in our understanding of zebrafish hematopoiesis do not allow us to fully resolve the nature of each observed switch. Our data support that both maturational and cellular switching processes contribute to the changes in the overall *globin* expression observed throughout development.

The functional identification of the LCR in the zebrafish through the use of DNase I hypersensitivity mapping and GATA1 ChIP demonstrates the power of this technique to quickly identify key hematopoietic regulatory sequences. The identification of novel co-localized peaks within mature red cells may indicate the importance of such regions. In addition, the comparison

of such dataset between red cell from multiple waves or maturational stages could provide insights into the dynamics of particular regions and their putative relevance to the functional changes occurring. These techniques are complementary to and not redundant with sequence analysis techniques as the rearrangement of elements can obscure sequence similarity despite functional similarity. In addition, regions identified by sequence similarity may not be identified by the genomic state of the region (MCS-E2) [199], and regions identified by sequence homology may not be as robust regulatory regions as those identified by genomic state [201].

The well documented [33,34] high level of syntenic conservation of the organization of the *globin* locus throughout evolution was able to both inform our work and is highly suggestive that there can be a fluid translation of information between zebrafish and mammalian systems. The sequence conservation observed through our analysis and annotation of both the coding and some non-coding regions of the loci concur with the assessment of previous work that both the synteny and regulatory regions have changed little over 500 million years of evolution [40]. In particular, the sequence conservation of the *cis*-regulatory elements, including GATA-1, NF-E2 and SCL binding motifs, suggest functional conservation in the *trans*-regulatory networks interacting with the *cis* elements and demonstrate the evolutionary constraints on this essential process. The syntenic conservation of flanking genes furthers this point, including *nprl3* which contains the LCR in mouse, humans, pufferfish, medaka [45,201] as well as the regulatory element identified here.

The identification of the functional LCR in zebrafish can facilitate the dissection of the functional role(s) of surrounding regions in the process of *globin* switching. With the relative ease of transgenics in the zebrafish, fragments of DNA can be coupled to GFP and analyzed *in vivo* in the context of wildtype or hematopoietic mutants or morphants. The transgenic construct

generated here recapitulates endogenous *globin* expression (Figure 2-5C, 2-1) and mimics the temporal expression changes observed for analogous constructs in the mammalian system [49,204]. Assessing the affects of alterations to the construct can, as shown here, be more easily tested in the zebrafish than a mammalian model. The loci's genomic context, presence of "intervening" sequence between the LCR and proximal promoter/coding sequence and the absence of additional genes in the region have all been investigated and/or shown to have an effect on *globin* switching [35,38,40,42,46,170,205]. Determining the role(s) of these additional regulatory elements can provide a better understanding of *globin* switching. For instance, it is known that the *zinfandel* mutation (*zin*), which has been linked to the major *globin* locus, alters embryonic *globin* expression from both loci [30]. This mutation is likely in a critical regulatory element, and identifying the mutation and its role in the regulatory network using this information and these tools will shed light on the broader process of *globin* switching.

Chapter 3

The Wnt and thyroid hormone receptor pathways modulate *globin* gene expression in zebrafish and human systems

Attributions

I designed and carried out the knockdown screen, including the microarray analysis and assembly of the candidate gene list. I also designed and carried out the follow up experiments in embryos and larva, including drug, morpholino, and combination treatments assayed by qPCR and blood smear and whole mount antibody staining, as well as the developmental time course expression assessments in wild type zebrafish. I identified and performed *in situ* hybridization to demonstrate the effect of T4 and XAV939 on adult globin expression. I also helped in the development of the adult globin specific zebrafish antibody. I designed and carried out the RNA-seq and red cell maker experiments in the zebrafish embryos. Analysis of the RNA-Seq data was carried out by Rory Kirchner and Oliver Hofmann and was supervised by Winston Hide. Rory Kirchner wrote the corresponding methods section. Yi Zhou assisted in the annotation of the results. In the adult zebrafish, I designed and carried out the dronedarone treatments experiments and the associated qPCR and antibody staining assays. In conjunction with Eirini Trompouki, I helped perform the ChIP-seq experiments in adult zebrafish red cells (RXR only), larval zebrafish red cells, and K562 cells. Eirini Trompouki received technical assistance from Sonja Boatman, and bioinformatic analysis was performed by Anthony DiDiase (adult zebrafish and K562 cells) and Zi Peng Fan (larval zebrafish cells). Zi Peng Fan also assisted in writing the corresponding methods section and was supervised by Richard Young. I received technical assistance in these efforts from Rebecca Maher, Elizabeth Riley, and Sruthi Satishchandran during different time periods. I prepared the manuscript, with the exception of the methods sections noted below, and prepared all figures for publication.

David Wiley designed and carried out the chemical screen with the technical assistance of Grant Rheingold. David Wiley also designed and preformed the K562 and $CD34^+$ cell

treatments experiments and assays, and wrote the corresponding methods. He also assisted in the development of the adult globin specific zebrafish antibody, and the associated whole mount antibody staining experiments at 24 and 48 hpf. He received technical assistance in these efforts from Sonja Boatman.

The adult globin specific zebrafish antibody was developed in conjunction with Mohan Brahmandam, Craig Bencsics, and Edward Greenfield the Dana-Farber Cancer Institute Monoclonal Antibody Core (Boston, MA). Edward Greenfield also assisted in writing the corresponding methods section.

Abstract

Hemoglobin switching is a highly regulated set of molecular changes that alter the protein components of the oxygen carrying molecule hemoglobin throughout development. The process has been studied since its discovery in the 19th century as a model of general gene regulation and because of its potential clinical implications for the treatment of the hemoglobinopathies. However, the process is still not completely understood. To identify new potential regulators of *globin* expression, we performed both a morpholino knockdown and a high throughput chemical screen for increased adult *globin* expression in 24 hpf zebrafish. We identified four genes and 52 chemicals that alter normal *globin* expression. The combined knockdown of two of the genes identified, *tcf7l2* and *ncoa1*, robustly and significantly increased the number of adult *globin* positive cells. Similarly, two chemicals identified, XAV939, a Wnt pathway inhibitor, and L-thyroxine (T4), thyroid hormone, individually and in combination significantly increased adult *globin* expression in embryos. Further analysis of the thyroid hormone pathway showed it specifically induced adult *globin* mRNA and protein without affecting general development or red cell production. Inhibition of the thyroid hormone receptor pathway in adult zebrafish increased embryonic and larval *globin* expression, which indicates the pathway plays a role in the maintenance of the adult state. T4 treatment of fetal globin-expressing K562 and differentiated human fetal CD34^{+} cells increased levels of adult β-globin, demonstrating the role of this pathway in globin switching is conserved in mammals. ChIP-seq analyses from both zebrafish and humans, taken together with the genetic and chemical biologic results, strongly suggest that thyroid hormone receptor activation directly alters the chromatin structure of the *globin* locus. My work defined novel roles for two pathways in *globin*

regulation, and provides strong support for these pathways to be investigated as clinical targets for hemoglobinopathy therapies.

Introduction

Globin is the protein component of hemoglobin, the critical oxygen carrying molecule responsible for supplying oxygen to the tissues of all vertebrates [24]. Hemoglobin switching is a critical developmental process during which the complement of *globin* genes that an organism expresses is altered by a tightly regulated molecular process to meet its changing needs during embryonic, fetal, and adult life [38,171,199]. Many upstream pathways that regulate red cell specification, development, and differentiation also influence the molecular environment in which *globin* expression occurs, but the process is ultimately regulated directly at *globin* loci. An upstream enhancer known as the locus control region (LCR) plays a particularly important role within the group of *cis*-regulatory elements modulating the expression of *globin* genes, and contains a number of DNase I hypersensitive sites that are essential to its proper function [38,98,206,207]. Hemoglobin and its regulation have been studied in a large range of organisms from single-celled species to humans [33], and significant conservation has been observed including between frogs [208], fish [34,209], mice, and humans [38]. Previous work from our lab has also demonstrated this conservation of *cis*- and *trans*-regulatory elements in the zebrafish [3,30,32,169], making the zebrafish a suitable system to further investigate *globin* switching.

The normal regulation of this process is disrupted in a collection of clinical disorders known as the hemoglobinopathies [210–212], which remain a major health concern worldwide [176,177]. For a number of these disorders, altering globin expression can helps to improve the clinical course of their disease. For example, increasing γ-globin, typically a fetal globin, expression in sickle cell patients can ameliorate many of the symptoms of their disease [79,176].

By better understanding the endogenous genes, pathways, and mechanisms that regulate developmental globin switching, it may illuminate additional potential ways to clinically treat one or more of these disorders.

The regulatory pathways that may directly influence hemoglobin switching in humans have been studied using a number of different model systems including cell culture [81,213], mice [10], transgenic mice [52] and humans [38]. Work in these systems has yielded significant progress towards a greater level of understanding, including the elucidation of the critical role of BCL11a in fetal globin suppression [63,64]. However, gaps in our understanding remain that are addressable in additional model systems. Previous work from our lab has helped to establish the zebrafish as another suitable model system for the study of conserved hemoglobin regulatory pathways [169]. This makes additional avenues of investigation available, such as large scale knockdown and chemical screening in a whole organism, which can facilitate additional discoveries [109,110].

The Wnt pathway is a large, complex signaling pathway critical to the normal development of a number of tissues in a number of species [139,141,214]. This includes essential functions in the hematopoietic system [142,143]. One of the many proteins involved in the Wnt pathway is the transcription factor *tcf7l2* (Tcf4), which, binds to and modulates the expression of Wnt target genes [137]. Despite this pathway's known role in general development, there is no specific role known for Wnt in *globin* gene switching or regulation.

The nuclear hormone receptor pathway is a large, complex superfamily with known critical roles in a wide range of tissues and contexts, including general organismal development [146]. Specifically, the thyroid hormone receptor pathway plays a role in the metamorphosis of

amphibians and teleosts [155,215], as well as postembryonic remodeling of some mammalian species such as mice and humans [159]. In some cases, such as that of the frog, this has been directly linked to globin switching [158]. However, for most systems the links between thyroid hormone and development as well as between development and *globin* switching are known, but the direct link between thyroid hormone and *globin* gene expression has not been examined. The thyroid hormone receptor pathway and the Wnt pathway interact in a number of systems and tissues [165–168], but this interaction has not been shown to influence *globin* gene expression, directly or indirectly.

Here, we demonstrate that both the Wnt and thyroid hormone receptor pathways are involved in the regulation of *globin* gene expression. In addition to the ability of both pathways to modulate expression independently, we demonstrate that the pathways work together to influence the *globin* expression pattern. In the zebrafish embryo, a specific increase of adult *globin* genes is observed with activation of the thyroid hormone pathway and inhibition of the Wnt pathway at the mRNA level and thyroid hormone pathway activation at the protein level. The particularly robust affect of thyroid hormone pathway activation also translates to zebrafish larva and human K562 and differentiated primary fetal $CD34^+$ cells. We have also demonstrated that thyroid hormone is necessary for the maintenance of adult *globin* expression in adult zebrafish. This has potential clinical implications for increasing HbF in humans and the treatment of hemoglobinopathies. ChIP-Seq analysis supports that this pathway is acting directly on the *globin* locus. NCOA1, RXRα, and TRβ are bound to a number of critical DNase I hypersensitive regions in the LCRs of both the α- and β-globin in human erythroid cells, and RXRα is bound to multiple bidirectional promoters in adult zebrafish peripheral blood and associated with key changes to H3K27ac within the *globin* clusters. These data clearly

demonstrate that both the Wnt and thyroid hormone receptor pathways regulate *globin* expression, with the thyroid hormone receptor acting directly on the locus to establish and maintain adult *globin* expression.

Materials and Methods

Zebrafish maintenance

Zebrafish were staged, raised, and maintained as previously described [188,189]. All zebrafish experiments and procedures were performed as approved by the Boston Children's Hospital institutional Animal Care and Use Committee.

Mircoarray analysis

Previously generated raw microarray data was obtained [216] and analyzed using the ArrayStar software package (DNASTAR, Madison, WI). When comparing data sets between waves (primitive, fetal definitive, adult definitive), all replicates and erythroid maturational stages (proerythroblasts, basophilic erythroblasts, orthochromatic/polychromatic erythroblasts, reticulocytes) were combined and treated as replicates. A p-value<0.01 and a fold change of at least 8 were used to initially define differentially regulated genes. Gene Ontology (GO) terms were used to identify transcription factors within the list of differentially regulated genes [217]. From these transcription factors, a 2 fold cut off was used to determine if gene expression changed between each pair of waves (increase, decrease, no change), from which a simplified expression pattern of the gene over developmental time was determined.

Morpholino injection

Morpholinos utilized in the knockdown screen were designed, synthesized (Gene-Tools, Philomath, OR) and injected as previously described [110] with the exception of Pbx1a, Jun, Lmx1b.1, Lmx1b.2, Prox1, and Hif1ab for which the designs have been previously described [218–222].

Chemical screening

The ICCB Biomol Bioactives (480 compounds), the NIH Clinical Collection (760 compounds), the Lopac library, and the Small Biomol libraries were screened. Zebrafish embryos were collected from specially designed large breading tanks [223]. The unfertilized and damaged embryos were sorted out and 10 embryos per well were added to 96 well mesh bottom plates submerged in E3. At 50% epiboly the plates were transferred to 96 well receiver plates containing 0.3% DMSO in E3 media with a 1:300 dilution of a chemical from the library. The embryos were pronase-treated, fixed, and processed for ISH at 24hpf. The effects of chemicals were scored under a bright-field stereo microscopy and a prioritized chemical list was generated for compounds that cause the most significant increases.

Drug treatments

T4 treatments were carried out using L-Thyroxine (T4; Sigma, St. Louis, MO). A master stock of 5 mM T4 dissolved in dimethyl sulfoxide (DMSO; Catalog #D5879, Sigma) and 150 uM NaOH was diluted in additional DMSO to generate the working stock, which was then diluted 1:100 in E3 and applied to the embryos or larva. An equal dilution of DMSO was used as the vehicle treatment. A master stock of 10 mM XAV939 (Catalog#3748, Tocris, Bristol, United Kingdom) dissolved in DMSO was diluted in additional DMSO to generate a working

stock, which was then diluted 1:100 in E3 and applied to the embryos or larva. An equal dilution of DMSO was used as the vehicle treatment.

Larval red cell collection

Larval peripheral blood was collected by cutting individual larva posterior to the yolk sac extension while the larva was immersed in 1X Dulbecco's phosphate buffered saline (Catalog # D8537, Life Technologies, Grand Island, NY) and heparin (1 USP/mL, Catalog #2106, Sigma) using 30G needles (Catalog #305106, Becton, Dickinson and Company, Franklin Lakes, NJ).

Dronedarone injections

A master stock of 50 mM dronedarone (Dron; Catalog #D9696, Sigma) dissolved in DMSO was diluted in additional DMSO to generate the working stock, which was then diluted in PBS to generate the injection dose with final concentrations of 5, 8, or 15 μM (30% DMSO final concentration). Injections were performed once per day for 4 days, approximately 24 hours apart, by retro-orbital injection [224] in alternating eyes. Peripheral blood was collected by heart puncture. Kidney marrow was collected by dissection followed by dissociation and filtering though a 40 μM filter.

Whole mount antibody staining

Whole mount fluorescent antibody staining was performed using standard procedures. Briefly, embryo were fixed in 4% paraformaldehyde, permeabilized with proteinase K (10 μg/mL in PBS), blocked in 10% lamb serum, incubated overnight at 4 °C with primary antibody, and incubated with a fluorescent secondary antibody. Samples were washed between each step with 0.8% Triton (Catalog #T8787, Sigma, St. Louis, MO) in PBS. Primary antibodies used

were the custom zebrafish adult globin antibody (1:500) and Anti-Hbbe-1.1 (NT), Z-Fish (1:500; Catalog #55608, Anaspec, Fremont, CA). Secondary antibodies used were Alexa Fluor® 488 Goat Anti-Mouse IgG (H+L) Antibody, highly cross-adsorbed (1:2000; Catalog #A11029, Life Technologies, Grand Island, NY) and Alexa Fluor® 594 Goat Anti-Rabbit IgG (H+L) Antibody (1:2000; Catalog #A11012, Life Technologies).

Blood smear antibody staining

Immunofluorescence on zebrafish peripheral blood was completed using standard protocols. Briefly, once cells were collected they were dried onto Poly-L-Lysine Coated Microscope Slides (Catalog #22247, Polysciences, Inc., Warrington, PA) in PBS. Samples were then fixed in 4% paraformaldehyde for 30 minutes, blocked in a 10% lamb serum/1% DMSO solution, incubated overnight at 4 °C with primary antibody, re-blocked, and incubated with a fluorescent secondary antibody. Samples were washed between each step with 0.1% Triton (Catalog #T8787, Sigma, St. Louis, MO) in PBS. 4',6-Diamidino-2-Phenylindole, Dilactate (DAPI; Catalog #D3571, Life Technologies, Grand Island, NY) at 1mg/L was also included prior to final washes in PBS to mark nuclei. Primary antibodies used were the custom zebrafish adult globin antibody (1:500) and Anti-Hbbe-1.1 (NT), Z-Fish (1:500; Catalog #55608, Anaspec, Fremont, CA). Secondary antibodies used were Alexa Fluor® 488 Goat Anti-Mouse IgG (H+L) Antibody, highly cross-adsorbed (1:2000; Catalog #A11029, Life Technologies, Grand Island, NY), Alexa Fluor® 594 Goat Anti-Rabbit IgG (H+L) Antibody (1:2000; Catalog #A11012, Life Technologies), Alexa Fluor® 568 Goat Anti-Mouse IgG (H+L) Antibody (1:2000; Catalog #A11004, Life Technologies, Grand Island, NY), and Alexa Fluor® 488 Goat Anti-Rabbit IgG (H+L) Antibody (1:2000; Catalog #A11008, Life Technologies). Images were captured using a

Nikon NiE C2Si upright confocal microscope with a 40x 0.8NA objective. Quantification was carried out using ImageJ software [225].

Cell culture

K562 cells were maintained in IMDM media with 10% Fetal bovine serum (FBS) and Pen/Strep. Human fetal CD34+ cells were obtained from Novogenix as frozen vials. These cells were expanded in StemSpan SFEM and CC100 cytokine (Stem Cell Technologies, Vancouver, Canada) for 6 days. After expansion, cells were differentiated toward the erythroid lineage using a previously described protocol [226].

For T4 treatments the serum was charcoal filtered to eliminate endogenous hormones. 3 g of activated charcoal (Catalog# C-5260, Sigma) was mixed with 100 ml of serum and incubated for 2hrs. The serum was then ultracentrifuged for 20min @ 25,000 rpm and the supernatant filtered through a 0.22 μm filter.

Monoclonal antibody generation

Female BALB/c, C56BL/6 and Swiss-Webster mice, 4–8 weeks old, were purchased from Charles River Laboratories (Wilmington, MA). The mice were initially immunized subcutaneously with 100 ug purified adult zebrafish peripheral blood lysate in Dulbecco's phosphate-buffered saline (PBS; Invitrogen, NY) and emulsified with an equal volume of complete Freund's adjuvant (Sigma Chemical Company, St Louis, MO). Two additional intraperitoneal immunizations of 50 ug were mixed 1:1 with incomplete Freund's adjuvant (Sigma Chemical Company, St Louis, MO) were given 2 and 4 weeks later. Ten days following the last immunization, a bleed was obtained by retro-orbital bleeding to check the antibody titer. Mouse #386-2 (C57BL/6) was intravenously boosted with 50 ug purified lysate 3 days prior

hybridoma fusion. The remaining two mice were given two additional boosts emulsified with an equal volume of Hunter's Titermax Gold adjuvant (Sigma Chemical Company, St Louis, MO). Another bleed was obtained 10 days after the last immunization and the titer checked. To further attempt to increase the titer, another two intraperitoneal immunizations, two weeks apart were given emulsified with an equal volume of Hunter's Titermax Gold + CpG (20 ug/mouse; InvivoGen, San Diego, CA) followed by two final intraperitoneal immunizations emulsified with an equal volume of incomplete Freund's adjuvant + CpG (20 ug/mouse). Mouse #386-3 (Swiss-Webster) was intravenously boosted with 50 ug purified lysate 3 days prior hybridoma fusion. All animals were acquired and maintained in accordance with the guidelines of the Institutional Animal Care and Use Committees of Harvard Medical School and the Guide for the Care and Use of Laboratory Animals (Institute of Laboratory Animal Resources, National Research Council).

The selected mice were rested for 4 weeks after the last immunization and then boosted by intravenous injection with 50 ug purified lysate in PBS. Four days later, the mice were sacrificed, and spleen and lymph node cell suspensions were prepared and washed with PBS. The cell suspensions were counted using a hemocytometer and mixed at a spleen:myeloma cell ratio of 2:1. The SP 2/0 myeloma cells (ATCC No. CRL8-006, Rockville, MD) used are incapable of secreting either heavy or light immunoglobulin chains [227]. Cells were fused using polyethylene glycol 1450 (ATCC) into eight 96-well tissue culture plates suspended in HAT selection medium according to standard procedures [228]. Clones were screened using a combination of western blotting and blood smear antibody staining using larval, adult, and mixed samples for comparison. A clone producing an IgG2a κ isoform specifically binding to adult zebrafish red cells was selected and used for antibody production.

RNA isolation and cDNA preparation

Pools of embryos or samples of tissue were collected and homogenized in Trizol reagent, Trizol LS reagent (Invitrogen, Carlsbad, CA), or using the RNEasy Plus Mini Kit (Qiagen, Germantown, MD). For Trizol and Trizol LS, total RNA was isolated using the manufacturer's protocol with the addition of GenElute LPA (Sigma-Aldrich, St. Louis, MO) as a carrier for small tissues samples. For RNEasy samples, total RNA was isolated using the manufacturer's protocol. For all samples, SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, Carlsbad, CA) was used to generate cDNA.

Quantitative real-time PCR

Quantitative real-time PCR (RT-PCR) was performed on a CFX384 Real-Time System with C1000™ Thermal Cycler (Bio-Rad, Hercules, CA) using SYBR Green Supermix (Bio-Rad). Zebrafish [169] and mammalian [69] *globin* primers have been previously described. The ProbeFinder online software tool (Roche Diagnostics, Basel, Switzerland) was used to design all other zebrafish RT-PCR primers used. Reactions were carried out in 5 μL; 2.5 μL of SYBR Green Supermix, 2 μL of diluted cDNA, and a final concentration of 200 nM of forward and reverse primers. Analysis was performed using the Bio-Rad CFX Manager software, version 3.1. For all analysis of zebrafish red cell samples, *slc4a1* was used as a control gene. For all mammalian samples, a combination of GAPDH and β-actin was used as a control. For all zebrafish whole embryo samples, *gapdh* was used.

In situ hybridization

The whole-mount *in situ* hybridization (ISH) protocol was carried out as previously described [197] using antisense probes amplified from digested plasmids. Scores for treatments

were determined by counting individual positive cells in the intermediate cell mass (ICM) of stained embryos at 24 hpf. Up to 39 individual cells were counted, after which the cells became too numerous to count. In initial treatments, these were classified and analyzed as "40." In subsequent treatments, levels above 39 cells were subjectively categorized into six categories, which were classified and analyzed as "40"-"45."

RNA-Seq

Samples were collected in Trizol reagent (Invitrogen, Carlsbad, CA), and RNA was purified according to the manufacturer's instructions. The RNA was further purified using the TURBO DNA-free Kit (Catalog #AM1907, Life Technologies, Grand Island, NY) and RNEasy Mini Kit (Catalog #74104, Qiagen,Valencia, CA). Samples were submitted to the Whitehead Institute Genome Technology Core for library construction and sequencing (Cambridge, MA).

To analyze the data Truseq adapter sequences and bases with quality scores < 20 were trimmed using cutadapt [229]. Trimmed reads were aligned using Tophat 2.0.5 [230] to build Zv9 of the zebrafish genome, using as annotations release 68 of the Ensembl gene annotations. A maximum of two mismatched bases per alignment was allowed. Reads mapping to genes were counted with htseq-count version 0.5.3 [231] and differential expression was called between groups using DESeq [232], with sharingMode="fit-only".

ChIP-Seq

ChIP- seq was performed as previously described [169] with the following modifications. For the H3K27ac ChIP-seq in zebrafish larval red cells, the three step lysis was modified such that the cells were diluted in solution 3 and sonicated. All samples were multiplexed using the NEBNext Multiplex Oligos for Illumina (Index Primers Set 1, E7335S, New England Biolabs,

Ipswich, MA) according to the manufacturer instructions. Analysis of the K562 data sets was performed as previously described [145].

Zebrafish ChIP-seq data sets were aligned using Bowtie (version 0.12.2) [191] to build version Zv9/danRer7 of the zebrafish genome with parameters "-k 2 -m 2 -n 2". MACS version 1.4.1 (Model-based analysis of ChIP-seq) [192] peak finding algorithm was used to identify regions of ChIP-seq enrichment over background. A *P*-value threshold of enrichment of 1e-9 was used for all data sets with parameters "--nomodel -g 1.4e+9 --keep-dup=2." To obtain the normalized read density of ChIP-seq data sets in any region, ChIP-seq reads aligning to each region were extended automatically by MACS and the density of reads per base pair (bp) was calculated. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per bp (r.p.m./bp). In order to determine the overlap between ChIP-seq datasets, a general linear model to identify regions with differential signal between different H3K27ac ChIP-seq profiles was adopted [233,234]. First, the set of genomic regions that were enriched for H3K27ac ChIP-seq signal in any one of the 3 data sets being considered were identified. Regions from a data set that overlapped with regions from another data set by 1 bp were merged together to form a representative region that spans the combined genomic region. For each region, the read density in reads per million per bp (r.p.m./bp) was calculated, and from this the relative read count of each region was obtained by multiplying read density by the length of the region. The edgeR package was used to model the variations due to differences in signal between two ChIP-seq data sets [233]. Sequencing depth and upper-quantile techniques were used to normalize all three data sets together before common dispersions were estimated. The statistical significance of differences between two ChIP-seq data sets was next calculated using an exact test and resulting

P values were subjected to Benjamini–Hochberg multiple testing correction (FDR). For robustness, only regions that were called "enriched" by MACS in one of the two datasets in the pair-wise comparisons were used for differential signal analyses. The final regions with differential signal were required to have an absolute log2 fold change of normalized H3K27ac signal greater or equal to 1 and FDR less or equal to 0.1.

All ChIP-seq results were visualized using the UCSC genome browser (http://genome.ucsc.edu).

DNase I Hypersensitivity

DNase I hypersensitivity analysis of K562 cells (wgEncodeUwDnaseSeqRawSignalRep1K562.wig.gz) was obtained from the ENCODE Consortium Data Coordination Center at UCSC [235] and visualized using the UCSC genome browser (http://genome.ucsc.edu). The data was produced by the laboratory of John Stamatoyannopoulous (University of Washington).

Results

Identification of regulators of *globin* gene expression by knockdown screen

In order to identify transcription factors that regulate *globin* expression, we tested whether *globin* expression could be altered by knocking down specific transcription factors with morpholinos in zebrafish embryos. Changes in α_{a1} expression were selected as the screening assay because baseline expression in the embryos is extremely low at 24 hour postfertilization (hpf) and significant increases would clearly indicate a disruption the in normal *globin* switching process. The adult *α-globin* gene α_{a1} is expressed at a low, nearly undetectable level in 24 hpf zebrafish embryos. By in situ hybridization (ISH), expression remains nearly undetectable until

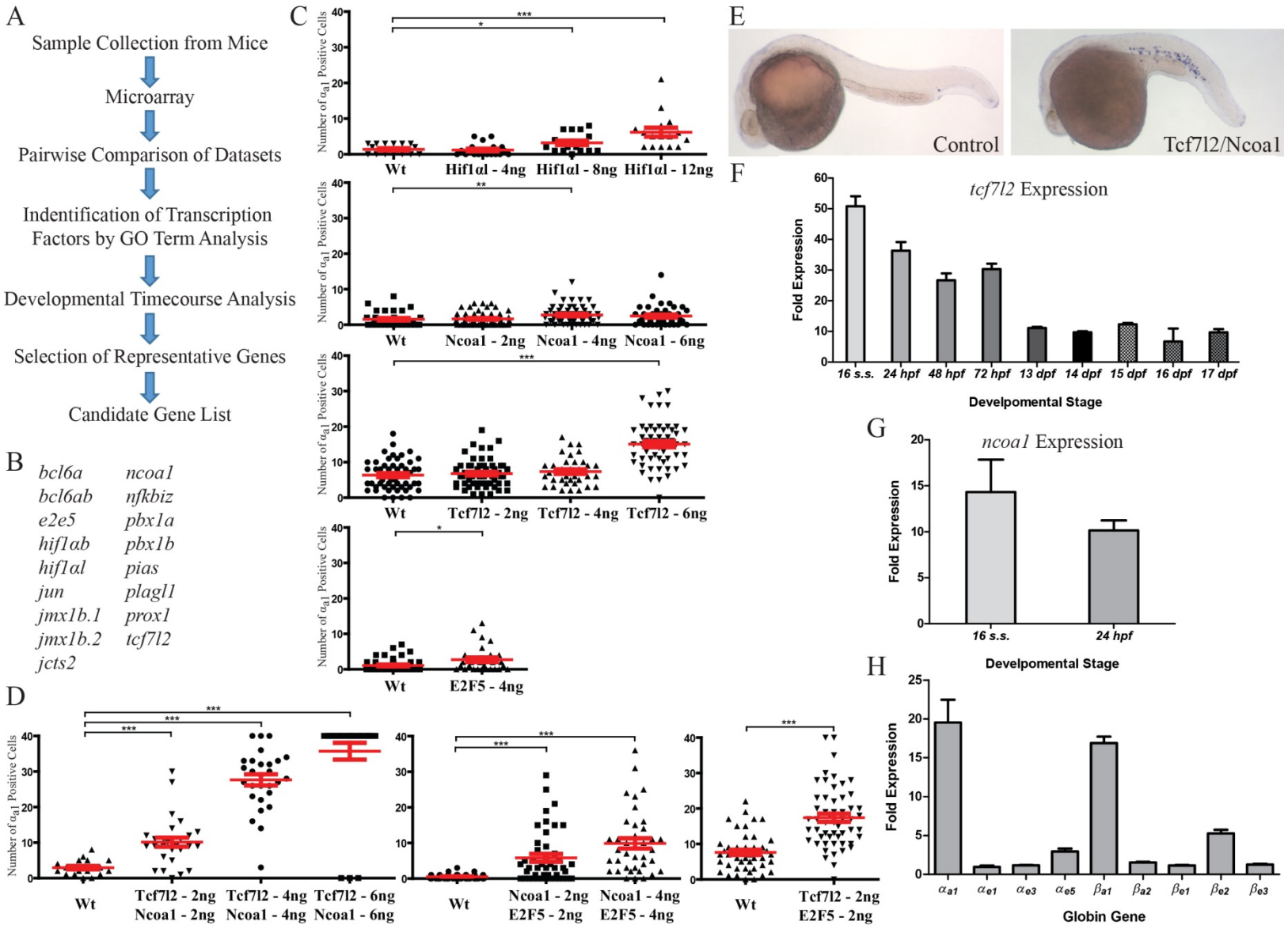
at least 5 days postfertilization (dpf), while quantitative real-time PCR (RT-PCR) analysis shows expression of α_{a1} does not surpass any other *α-globin* gene expression levels until 18 dpf [169]. After knockdown of individual transcription factors, α_{a1} expression was assessed by ISH at 24 hpf. If loss of a factor resulted in the disruption of the process, it would indicate the gene is involved in its normal regulation. By testing targeted, yet unbiased, genes by knockdown, we could potentially identify novel regulators of *globin* gene switching. Screening for, and characterizing, these factors in the context of a whole organism will help facilitate the translation of the pathways to other system such as mice and potentially humans. In addition, any genes identified would be potential repressors of adult *globin* expression, which could be useful during the erythroid differentiation of iPS and ES cells to promote an adult-like phenotype.

Candidate genes to be assessed were identified through bioinformatic analysis of previously generated murine microarray datasets [216] comparing distinct stages of red cell development (Figure 3-1A). Microarrays were performed on proerythroblasts, basophilic erythroblasts, orthochromatic/polychromatic erythroblasts, and reticulocytes isolated from murine yolk sac (YS), fetal liver (FL), and bone marrow (BM). YS and BM represent distinct red cell populations with distinct globin gene expression profiles, separated by the intervening FL stage [7,171]. All differentiation stages harvested from the YS, FL, and BM were combined to generate the Primitive (P), Fetal (F) and Adult (A) sample sets respectively. Pairwise comparison of these sets was performed and all differentially expressed genes identified. Gene ontology (GO) terms were used to identify the transcription factors among the genes identified. An analysis of the expression pattern of these genes through the three stages (P, F and A) was performed by comparing the fold changes identified in the pairwise comparisons. In order for a morpholino knockdown to more closely recapitulate the adult state and more likely induce adult

Figure 3-1. The combined knockdown of *tcf7l2* and *ncoa1* specifically increase adult *globin* expression in the embryo.

(A) Schematic for generation of the candidate gene list. The sample collection from mice and microarray were performed previously [216]. (B) Candidate genes tested by knockdown in the zebrafish. Morpholino screen results for (C) single morpholinos and (D) morpholino pairs. Each dot represents an individual embryo. The sensitivity limit of the assay is 39 positive cells per embryo, and embryos with 40 cells or greater were classified and analyzed as having 40 cells. All morpholino and morpholino combinations were injected at 3 doses; absence of a dose indicates that the level of toxicity/death prevented analysis. All embryos directly compared are from the same, single-pair, clutch. *p<0.05, **p<0.01and ***p<0.001. (E) ISH for α_{a1} in a 24 hpf control and Tcf7l2/Ncoa1 double morphant embryo. (F) Expression of *tcf7l2* over developmental time. All time points were normalized to the adult stage. (G) Expression of *ncoa1* over developmental time. All time points were normalized to 17 dpf. (H) Expression of *globin* genes in Tcf7l2/Ncoa1 double morphants (6 ng of each morpholino). The sample was normalized to wild type, uninjected clutch control embryos.

Figure 3-1. (Continued)



globin gene expression, representative transcription factors with lower expression levels in the adult state (e.g. reduced expression in A when compared to P and F) were further selected to generate a candidate gene list of 14 genes.

As a whole genome duplication event occurred in the zebrafish evolutionary lineage [114], zebrafish often possess multiple orthologs to a mammalian gene. The 14 mouse genes selected were found to have 17 orthologs in the zebrafish genome (Figure 3-1B). Each gene was knocked down individually by morpholino and the number of α_{a1} expressing cells in the intermediate cell mass at 24 hpf was counted in each embryo. Of the genes tested, four, *hif1al*, e*2f5*, *ncoa1* and *tcf7l2*, significantly increased the number of α_{a1} expressing cells over control embryos (Figure 3-1C). Pairwise knockdown of these four genes was then assessed. Of the six pairwise combinations, *ncoa1* with *e2f5*, *tcf7l2* with *e2f5* and *tcf7l2* with *ncoa1* produced greater increases in α_{a1} expression (Figure 3-1D). The largest increase in the number of α_{a1} expressing cells was obtained with the combined knockdown of *tcf7l2* and *ncoa1*. Many embryos injected with the highest dose of morpholino (6ng of each), and some injected with the second highest dose (4ng of each), had more α_{a1} expressing cells than were readily countable (above 39 cells per embryo) and recorded and calculated as "40" (Figure 3-1D, E). This significant increase demonstrates that both *tcf7l2* and *ncoa1* regulate *globin* gene expression.

The expression of *tcf7l2* and *ncoa1* were also assessed in the zebrafish. The expression of *tcf7l2* is over 50 fold higher in 16 somite stage (s.s.) embryos than in adults, and its expression gradually decreased between these stages (Figure 3-1F). In peripheral blood cells, expression is also approximately 1.5 fold higher in 16 dpf larva when compared to adult red cells (data not shown). *Ncoa1* expression was undetectable in most stages after 24 hpf, including larval and adult peripheral blood, but was over 14 fold higher in 16 s.s. embryos than 17 dpf larva (Figure

3-1G). Overall, these expression patterns indicate that *tcf7l2* and *ncoa1* are preferentially expressed in the developing embryo compared to the adult, which is consistent with both the microarray data from mouse erythroid cells and the role of these factors as repressors of adult *globin*.

In order to determine the effect of this knockdown on the complete *globin* profile of the embryos, we performed RT-PCR on the double morphants for the *globin* genes. As expected from the results obtained by ISH, the expression of α_{a1} increased, as compared to wildtype. α_{a1} expression increased over 19.5 fold, and the other major adult *globin*, β_{a1}, increased over 16.5 fold (Figure 3-1H). The minor adult *globin* β_{a2} and larval *globin* genes β_{e2} and α_{e5} were also modestly increased at over 1.5 fold, 5 fold and 2.5 fold respectively. The other embryonic/larval *globins* (α_{e1}, α_{e3}, and β_{e1}) maintained similar expression levels to controls, and the embryonic *globin* β_{e3} was only increased by about 1.2 fold (Figure 3-1H). This specific increase in the adult *globin* genes demonstrates that knockdown of *tcf7l2* and *ncoa1* is affecting the expression of specific *globin* genes and therefore *globin* switching rather than global *globin* or red cell functions. While slight changes to the overall zebrafish morphology are observed at the highest doses of morpholino, significant increases in adult *globin* are observed at all combined doses, including those below which any obvious morphological disruptions were detected. This indicates the morphological disruptions observed at higher doses are secondary to the phenotype or due to morpholino toxicity, and the increase is not an indirect affect of the morphological changes.

Identification of regulators of *globin* gene expression by a chemical screen

A chemical screen was also performed to identify potential regulators of *globin* switching. Embryos were collected, treated, and assessed using the same ISH assay utilized in the morpholino screen (Figure 3-2A). Of the 3120 chemicals tested, 52 (~1.67%) were identified as preliminary hits that increase adult *globin* expression, while the remaining approximately 98% did not affect the number of α_{a1} expressing cells and/or substantially affected the morphology of the treated embryos. One of the chemicals identified was L-thyroxine (Figure 3-2B). L-thyroxine (T4) is one form of thyroid hormone, and significantly increased the number of number of α_{a1} positive cells in a dose responsive manner, with many of the embryos treated at the highest dose exceeding the upper limit of the scale (Figure 3-2C). The chemical XAV939, which was not included in the chemicals used in the screen, was identified in a supplemental screen carried out in the same manner. XAV939, a tankyrase inhibitor that inhibits the Wnt pathway [236], also significantly increases adult *globin* expression in a dose responsive manner (Figure 3-2C).

***Tcf7l2* and *ncoa1* act to regulate *globin* gene switching through the Wnt and thyroid hormone receptor pathways respectively**

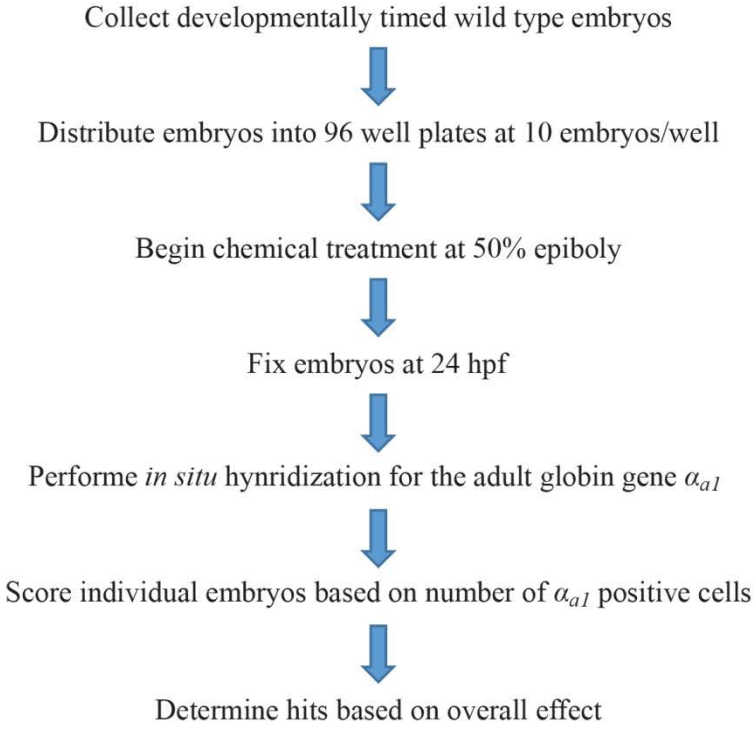
With the identification of two pathways, Wnt and thyroid hormone, and two genes, *tcf7l2* and *ncoa1*, we wanted to determine the relationship(s) between these four elements. As *tcf7l2* and *ncoa1* are associated with the Wnt and thyroid hormone pathways respectively, we hypothesized that *tcf7l2* regulates *globin* gene expression though the Wnt pathway and *ncoa1* regulates *globin* expression though the thyroid hormone pathway.

Figure 3-2. Activation of the thyroid hormone receptor pathway or inhibitor of the Wnt pathway leads to increase in adult *globin* expression in the embryo.

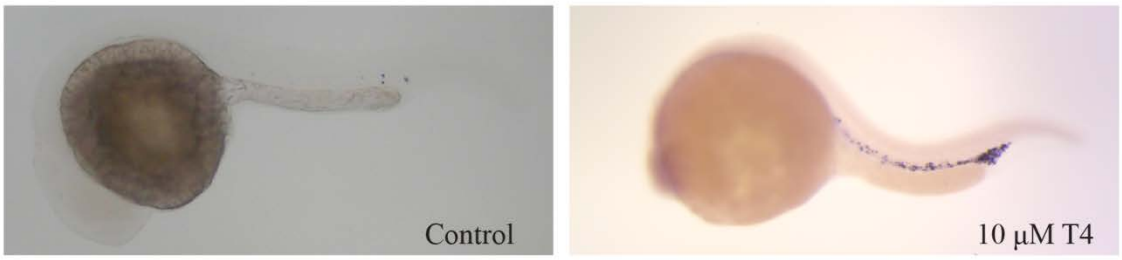
(A) Schematic for the chemical screen. (B) α_{a1} *in situ* for a control and T4 treated embryo at 24 hpf. (C) Chemical treatment results for T4 and XAV939. Each dot represents an individual embryo. The sensitivity limit of the assay is 39 positive cells per embryo, and embryos with 40 cells or greater subjectively categorized into six categories, which were classified and analyzed as "40"-"45." All embryos directly compared are from the same, single-pair, clutch. *$p<0.05$, **$p<0.01$and ***$p<0.001$.

Figure 3-2. (Continued)

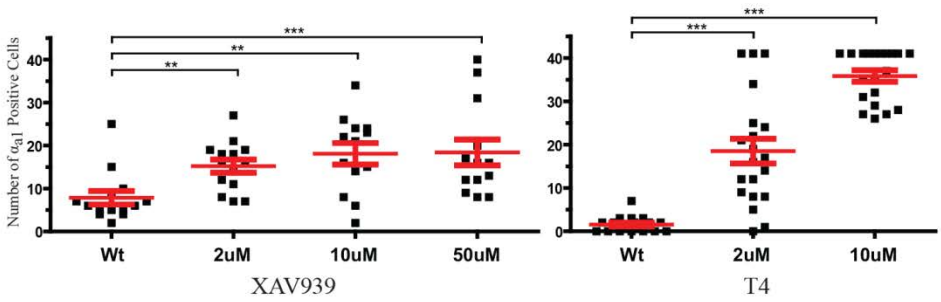
A



B



C



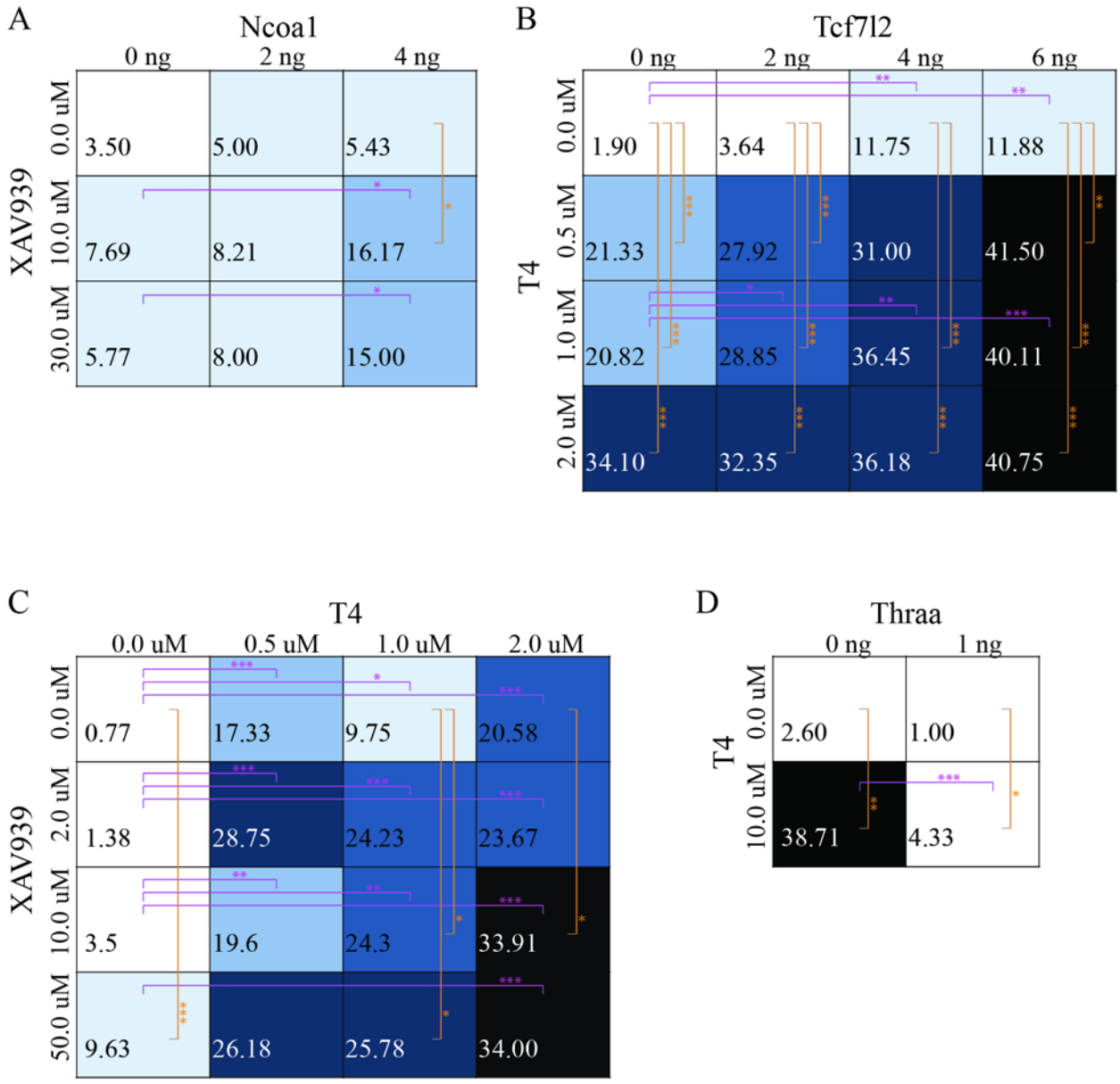
As the combined knockdown of *tcf7l2* and *ncoa1* further increase adult *globin* expression in the embryo, substitution of XAV939 or T4 respectively should similarly work together. As expected from the morpholino and chemical screen results, the knockdown of *tcf7l2* and *ncoa1* individually increased the number of α_{a1} expressing cells, as did treatment with T4 and XAV939 individually. The Wnt pathway inhibitor XAV939 and knockdown of the thyroid hormone associated coregulator *ncoa1* resulted in an additional increase in the number of α_{a1} expressing cells for the combined treatments of 10 μM/4 ng (Figure 3-3A). An increase in the number of α_{a1} expressing cells was also observed for the thyroid hormone pathway agonist T4 and knockdown of the Wnt pathway transcription factor *tcf7l2*. A number of dose combinations led to increases more significant than the corresponding doses of the individual agents. For example, the combined dose of 1.0 μM of T4 and 6 ng of the *tcf7l2* morpholino yielded an average of 40.11 α_{a1} positive cells per embryos, significantly more than the 20.82 and 11.88 observed for single agents respectively (Figure 3-3B). Similarly, co-treatment with XAV939 and T4 should produce a greater increase in the number of α_{a1} expressing cells. This was also observed at a number of co-treatments doses, such as 10.0 μM XAV939 and 2.0 μM T4. With this combined treatment, the average number of positive cells per embryos reached 33.91, significantly more than the 3.5 and 20.58 observed with the individual drugs (Figure 3-3C). Together, these results demonstrate that *tcf7l2* and *ncoa1* are functioning as members of the Wnt and thyroid hormone receptor pathways respectively in their role regulating *globin* switching. They also support that these pathways are functioning in parallel, and therefore can be modulated in conjunction to achieve greater changes in *globin* gene expression.

The zebrafish thyroid hormone receptors, *thraa*, *thrab*, and *thrb*, were knocked down in control embryos as well as prior to T4 treatment. *Thraa* and *thrb* are maternally deposited and

Figure 3-3. The regulation of *globin* gene expression by *tcf7l2* and *ncoa1* are through the canonical Wnt and thyroid hormone receptor pathways.

Co-treatments with (A) Ncoa1 morpholino and XAV939, (B) Tcf7l2 morpholino and T4, (C) T4 and XAV939, and (D) Thraa morpholino and T4. Each column or row was treated with the listed morpholino or drug, and each box shows the average number of α_{a1} positive cells in the ICM at 24 hpf, using the scale that includes the categories 41-45. The background shade of the boxes reflects the average number of adult positive cells (the darker the higher). Significance is indicated by the horizontal (purple) and vertical (orange) bars. *$p<0.05$, **$p<0.01$ and ***$p<0.001$.

Figure 3-3. (Continued)



expressed in the developing zebrafish embryo during the early treatment time window [237,238], so were potential candidates for transmitting the ligand signal into gene expression changes. Knockdown of Thraa blocked the induction of abnormal adult *globin* positive cells by T4 (Figure 3-3D). This result was confirmed with a second morpholino (data not shown). This demonstrates that, as expected, T4 is signaling thought the thyroid hormone receptor, most likely Thraa, to modulate *globin* gene expression.

The ability of both Wnt pathway modulators, XAV939 and *tcf7l2*, and both thyroid hormone pathway members, T4 and *ncoa1*, to function interchangeably in their ability to regulate *globin* gene expression clearly demonstrates that the genes identified are functioning thought the Wnt and thyroid hormone pathways to modulate *globin* gene expression. Our studies confirm the roles of the thyroid hormone and Wnt pathways, both independently and in combination, in the modulation of adult *globin* gene expression.

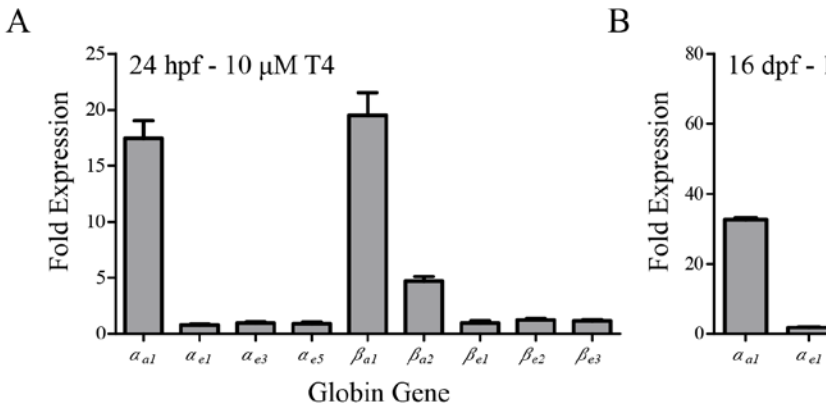
Activation of the thyroid hormone receptor pathway specifically increases adult *globin* gene expression

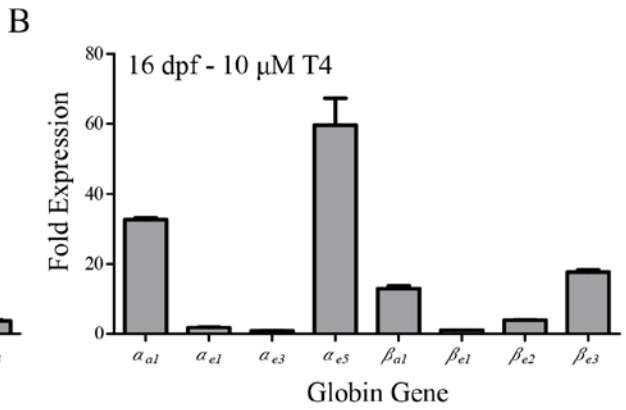
In order to determine the effect of T4 treatment on the complete *globin* profile of embryos at 24 hpf, RT-PCR for the *globin* genes was performed. All of the adult *globins*, including the expected increase in α_{a1}, were increased over control (Figure 3-4A). The largest increase was observed for β_{a1} (~19.5 fold), the second most for α_{a1} (~17.46 fold), and an approximately 4.5 fold increase was observed for the minor adult *globin* β_{a2}. The increase in α_{a1} and β_{a1} was comparable to that observed in the double morphant (Tcf7l2/Ncoa1), but the increase in β_{a2} genes is substantially greater (compared to ~2 fold; Figure 3-1H). The larval *globin* β_{e2} is also modestly increased compared to the control, but the larval α-*globin* α_{e5} is

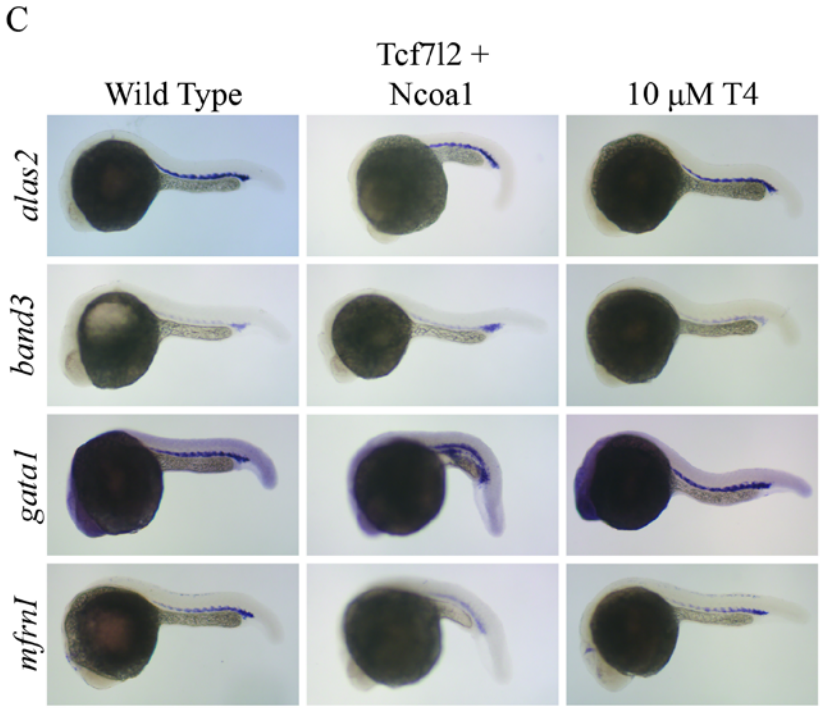
Figure 3-4. T4 specifically increases adult *globin* gene expression.

Expression of *globin* genes in T4 treated (A) 24 hpf embryos and (B) 16 dpf larva. The sample was normalized to vehicle treated control embryos. (C) *In situ* hybridization for red cell markers in wild type control, Tcf7l2/Ncoa1 double morphant, and T4 treated embryos at 24 hpf. (D) All genes significantly altered (padj<0.1) in T4 treated (10 μM) 24 hpf embryos as assessed by RNA-seq. Positive fold changes are increases and negative fold changes are decreases compared to vehicle treated control embryos.

Figure 3-4. (Continued)







RNA-Seq
24 hpf - 10 µM T4

D

Gene Symbol	Fold Change (Log2)	padj
hbaa1	5.1582	0.0004
hbaa2	3.2513	0.0077
hbba2	2.0796	0.0001
hsd11b2	1.4690	0.0004
krt42	0.6153	0.0818
fasnl	-0.8794	0.0000
crybb1	-0.9117	0.0821
"novel miRNA"	-0.9364	0.0000
grk5	-0.9527	0.0772
mucin13	-1.0601	0.0031
plekhf1	-1.1791	0.0094
LOC799595	-1.2059	0.0185
zgc:163121	-1.2080	0.0000
es1	-1.6675	0.0000
wu:fj08f03	-1.9527	0.0000
gif1	-2.2549	0.0615
fosl1	-2.5178	0.0000
slc16a9a	-2.6910	0.0068
irg1l	-3.5195	0.0000
sult5a1	-4.7784	0.0000
zgc:195209	-5.8090	0.0000

modestly decreased. The embryonic/larval *globins* (α_{e1}, α_{e3}, and β_{e1}) are slightly decreased compared to control. The embryonic *globin* β_{e3} is increased by approximately 1.15 fold (Figure 3-4A). The preferential increase in adult *globin* genes, and even modest decrease in some embryonic/larval *globins*, strongly indicates a specific effect on the regulation of the endogenous *globin* switching process.

Expression of the *globin* genes in zebrafish embryos begins at approximately the 12 somite stage [30], so T4 treatments beginning at 50% epiboly can begin to affect *globin* expression prior to the establishment of the normal *globin* gene complement. In order to test whether T4 treatment can disrupt a normal, established *globin* expression pattern, we treated zebrafish larval (15 dpf) and compared the *globin* expression to wild type controls (Figure 3-4B). While a similar pattern of increase is observed at 15 dpf as was at 24 hpf, the increase in α_{a1} expression is greater (over 32 fold) and the increase in β_{a1} is less substantial (~13 fold). In addition, a marked increase in the embryonic *globin βe3* is also observed (Figure 3-4B), which is absent at 24 hpf (Figure 3-1H). This demonstrates that T4 is able to not only establish an altered *globin* expression profile, but supersede an existing, established expression profile. The ability to alter the *globin* expression profile of older embryos with larger numbers of red cells per embryo also facilitates further experiments.

In order to determine if red cells were otherwise impacted by either T4 or *tcf7l2* and *ncoa1* knockdown, critical red cell markers were assessed by ISH and compared to wild type controls at 24 hpf. The red cells specific genes *alas2*, *band3*, *gata1*, and *mfrnI* were all unaffected by either treatment shown to significantly alter *globin* expression (Figure 3-4C). Global expression changes resulting from T4 treatment were also assessed using RNA-Seq performed on treated 24 hpf embryos and vehicle treated controls. A limited number of genes

were identified that are significantly altered by *in vivo* T4 treatment (5 upregulated; 16 downregulated; Figure 3-4D). The largest increase observed is an over 35 fold increase in α_{a1} gene expression, an even greater fold increase than observed with T4 treatment at 24 hpf by RT-PCR (Figure 3-4A). The two next most significantly increased genes are the two minor adult *globin* genes with between approximately 4 and 10 fold increases (Figure 3-4D). The largest fold decreases observed are in zgc:195209, a chemokine ligand (~56 fold decrease), *sult5a1* (sulfotransferase family 5A, member 1; ~27 fold decrease), and *irg1l* (immunoresponsive gene 1, like; ~11.5 fold decrease; Figure 3-4D). It is also important to note that critical red cell regulators (i.e. *gata1*) and functional proteins (i.e. *band3*, *alas2*, *mfrn1*) were not identified as being misregulated by RNA-Seq (Figure 3-4D). Together, these data indicate that red cells in T4 treated embryos are healthy, functional, and present in normal numbers. Additionally, the extremely limited number of genes altered indicates that T4 treatment may be inducing a targeted, direct response that does not require the transcription of additional genes to mediate an upregulation of adult *globin* gene expression.

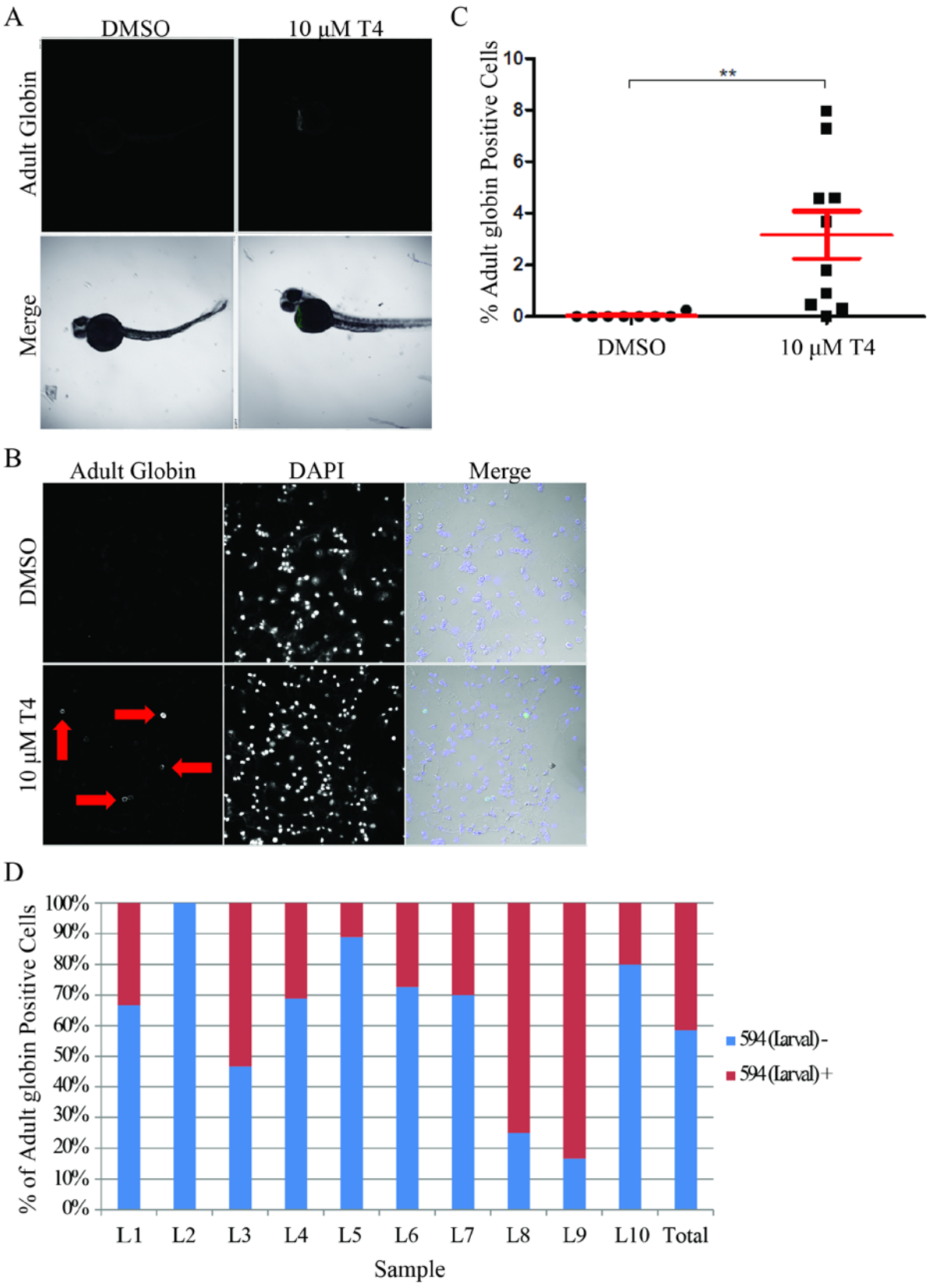
T4 treatment induces adult *globin* protein in a heterocellular manner

The specific upregulation of adult *globin*, as observed thought ISH, RT-PCR, and RNA-Seq, by T4 treatment demonstrates a clear increase in adult *globin* mRNA. In order to demonstrate that activation of the thyroid hormone receptor pathway also leads to an increase in adult globin protein, a monoclonal antibody was developed specifically against adult zebrafish hemoglobin. Embryos assayed at 24 hpf after T4 treatment did not express additional adult globin protein over control embryos (data not shown). However, embryos treated at the same stage expressed substantially more adult globin protein than vehicle treated controls at 48 hpf (Figure 3-5A). βe1 expression was also tested and appeared present and normal in both T4 and

Figure 3-5. T4 induces adult globin protein in a heterocellular manner.

(A) Whole mount antibody staining for adult globin. (B) Blood smear of larval peripheral blood stained for adult globin. (C) Quantification of larval peripheral blood adult globin antibody stains. The percent of positive cells was calculated using the number of adult globin positive cells and the total number of cells as determined by DAPI staining. Each dot represents an individual larva. **$p<0.01$. (D) Analysis of costaining of adult globin and the embryonic/larval globin β_{e1}. Each bar represents the sample from one larva ("L*") or the total for the group ("Total"). Adult globin positive cells from each larva were identified and then assessed for larval globin status. Each bar shows the percent of these adult globin positive cells that were larval globin negative (blue) and larval globin positive (red).

Figure 3-5. (Continued)



vehicle treated embryos at both 24 hpf and 48 hpf. In order to quantify the increase in protein, 15 dpf embryos were treated with T4 and lethally bled at 16 dpf. The peripheral blood was fluorescently stained using the adult specific globin antibody, and the number of peripheral cells positive for adult globin was compared to the total number of cells and the percent of cells positive for adult globin were calculated for each embryo (Figure 3-5B). The average percent of cells that were adult globin positive in the T4 treated embryos was 3.16% (n=10 embryos), significantly more than observed in the vehicle treated control (0.03%, n=8, $p<0.01$, Figure 3-5C). The ability to observe upregulation of adult globin in a subset of cells further supports the role of the thyroid hormone receptor pathway in the regulation of *globin* gene expression, and that no post-translational mechanisms prevent the translation of aberrant, T4-induced adult *globin* mRNA from being translated into protein.

In order to determine whether the larval cells transcribing and translating adult *globin* mRNA and protein, respectively, which typically express almost excursively larval globins, also expressed larval globin, we co-stained with an antibody against the embryonic/larval globin β_{e1}. We identified adult globin expressing cells isolated from T4 treated larva and determined whether they were co-positive for β_{e1}. Of the 193 adult globin positive cells observed from 10 larva, 58.55% (113) were single-positive (Figure 3-5D). In individual embryos with more than 10 cells observed (n=7), the average percent of the cells that were single-positive was 61.48% (Figure 3-5D). This demonstrates that the majority of T4 induced, adult globin protein expressing cells are not expressing the normal globin complement for their developmental stage in addition to adult globin. These data demonstrates that, as expected, aberrant adult *globin* mRNA expression precedes aberrant adult globin protein expression, which becomes detectable approximately 24 hours after T4 treatment begins.

The thyroid hormone pathway regulates *globin* expression in adult zebrafish

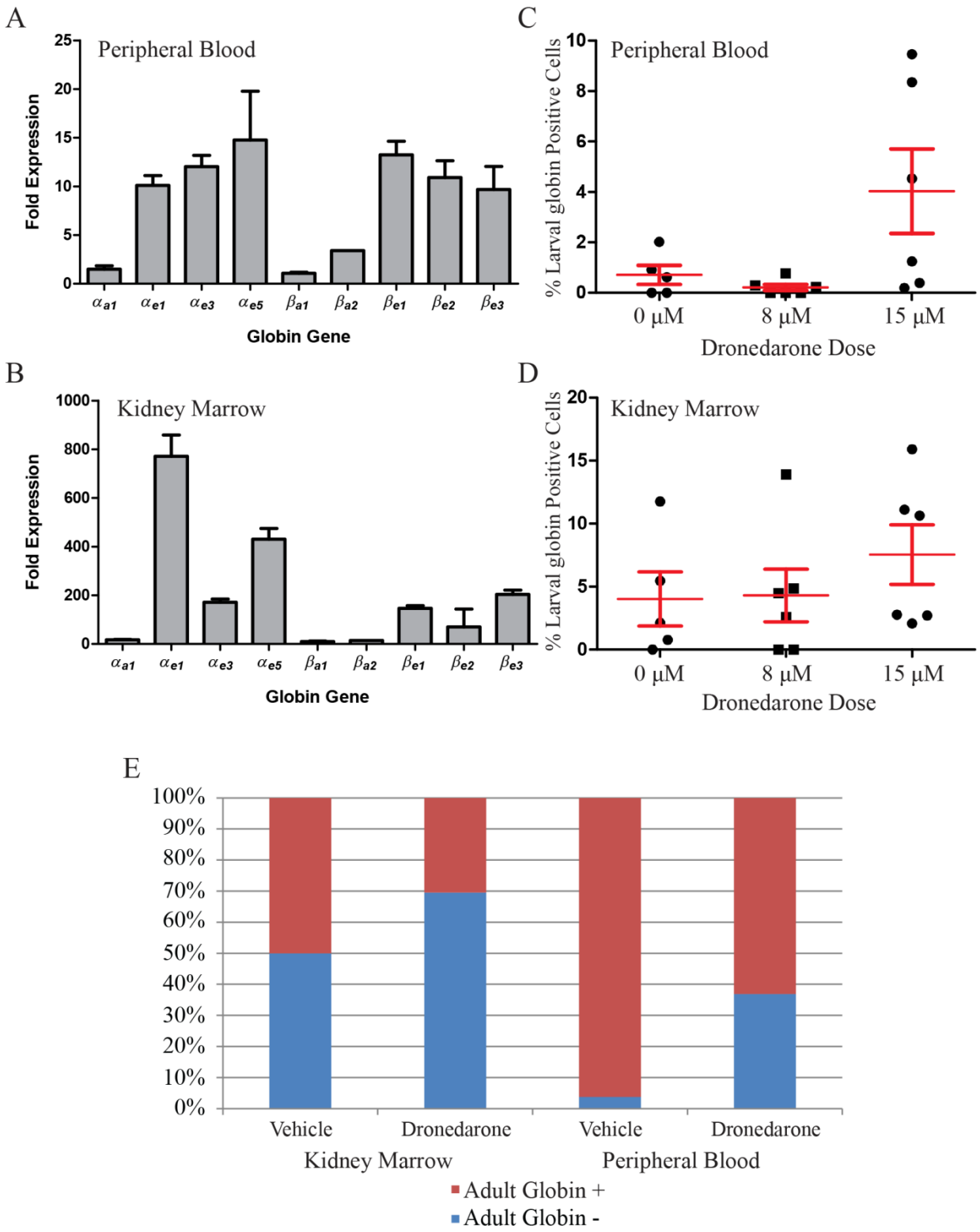
With the determination that the thyroid hormone receptor pathway regulates *globin* gene expression in the zebrafish embryo and larva, the role of the pathway in the adult context remained unclear. We hypothesized that, as agonism of the pathway upregulates adult *globin*, antagonism of the pathway would increase the expression of larval, embryonic/larval, and/or embryonic *globin* genes. The thyroid hormone receptor antagonist dronedarone (Dron) was used to inhibit thyroid hormone receptor function [239] in adult zebrafish. Dron was injected once daily for 4 days retro-orbitally, followed by collection of peripheral blood and kidney marrow on the fifth day. RT-PCR was used to assess the expression of the *globin* genes in Dron treated versus vehicle treated samples. In peripheral blood, the embryonic *globin* β_{e3} was upregulated over 9.5 fold over control (Figure 3-6A). The embryonic/larval and larval *globin* genes were also upregulated between 10 and 15 fold. The major adult *globins* α_{a1} and β_{a1} were largely unaffected (Figure 3-6A). In the kidney marrow, a similar pattern is observed, but with substantially larger increases. The embryonic, embryonic/larval, and larval *globins* increase between 70 and ~800 fold and the adult *globin* genes increase between 9 and 18 fold (Figure 3-6B). These results show that inhibition of the thyroid hormone pathway results in the upregulation of embryonic and larval *globin* in the adult kidney marrow and peripheral blood. As modulation of this pathway in the embryo as well as the adult leads to changes in *globin* gene expression, the thyroid hormone receptor pathway must play a role in both the establishment and maintenance of adult *globin* expression at the mRNA level.

Upregulation of larval *globin* expression at the protein level was assessed using fluorescent antibody staining against the embryonic/larval globin β_{e1}. After treatment with Dron, fish were individually bled and dissected to collect kidney marrow. As with the assessment of

Figure 3-6. Thyroid hormone regulates *globin* gene expression in the adult.

Expression of *globin* genes in dronedarone (Dron) treated (A) peripheral blood and (B) kidney marrow. The samples were normalized to tissues from vehicle treated controls. Quantification of larval globin positive cells in (C) peripheral blood and (D) kidney marrow or Dron treated adults. The percent of positive cells was calculated using the number of larval globin positive cells and the total number of cells as determined by DAPI staining. Each dot represents an individual. (E) Analysis of costaining of adult globin and the embryonic/larval globin β_{e1}. Each bar represents the cells from a single condition and tissue. Larval globin positive cells from each adult were identified and then assessed for adult globin status. Each bar shows the percent of these larval globin positive cells that were adult globin negative (blue) and adult globin positive (red).

Figure 3-6. (Continued)



the presence of adult *globin* protein in larval red cells, DAPI was used to determine the total number of cells present and the percent of β_{e1} positive cells was calculated for each animal. The average percent of β_{e1} positive kidney marrow cells in the Dron group was 7.54%, more than the 4.02% observed in the control group (Figure 3-6C). Similarly, the peripheral blood from the animals in the Dron group had more β_{e1} positive cells on average than the control group (4.03% versus 0.71%; Figure 3-6D). While neither increase is significant, this may be due to only a fraction of the individuals responding to the treatment, as the distribution in the treated groups is high, and/or the high level of variability of all samples. These experiments are being repeated with larger numbers to determine if the increase is significant.

Coexpression of adult globin protein was also assessed in these aberrant, thyroid hormone receptor antagonist induced cells. Of the 27 β_{e1} positive cells observed in the untreated peripheral blood samples, 96.3% were also adult globin positive, as compared to 63.16% in the Dron treated samples (n=19; Figure 3-6E). In the kidney marrow cells, the control cells were 50% co-positive (n=48) while the Dron samples was 30.43% co-positive (n=69; Figure 3-6E). The expression of the embryonic/larval globin in the adult stage in the untreated group suggests that, analogous to mammalian F cells, erythroid cells expressing globin genes from earlier developmental stages persist into adulthood. The higher percent of these β_{e1} expressing cells in the kidney marrow as compared to peripheral blood may suggest that less mature red cell progenitors may have a different globin expression pattern than peripheral blood. The ability of Dron to increase the percent of β_{e1} positive cells in the kidney marrow to a greater degree than in the peripheral blood may due to this higher basal expression. Induction of an embryonic/larval globin protein in the adult clearly demonstrates that the thyroid hormone pathway not only plays

a role in executing the normal developmental *globin* switching process, but also in the maintenance of the final, adult state.

The thyroid hormone receptor pathway regulates globin gene expression in the mammalian system

Both K562 cells, which express mainly fetal globin and a small amount of embryonic globin [81], and differentiating primary human fetal $CD34^+$ cells were treated with T4. As both cell types are analogous to the larval zebrafish globin state, we hypothesized that treatment would lead to an increase in adult *globin* expression as measured by adult β-globin. An increase in β-globin was observed in both K562 and $CD34^+$ cells after 24 and 72 hour treatments respectively (Figure 3-7). The dose responsive increases in adult β-globin in K562 cells ranged from 12.5 to 18.5 fold while they ranged from 30.5 to 38.5 in differentiated $CD34^+$ cells. The embryonic and fetal β-like globins only increased by just over one up to two fold, but increased by between 1 and 14 fold in the CD34+ cells, with the largest increases observed for embryonic globin (Figure 3-7). These results are comparable to or exceed those observed in whole zebrafish treatments (Figure 3-1H, 3-4A, B), demonstrating that the thyroid hormone pathway is both responsive in mammalian erythrocytes and regulates globin gene expression in a similar manner observed in the zebrafish system.

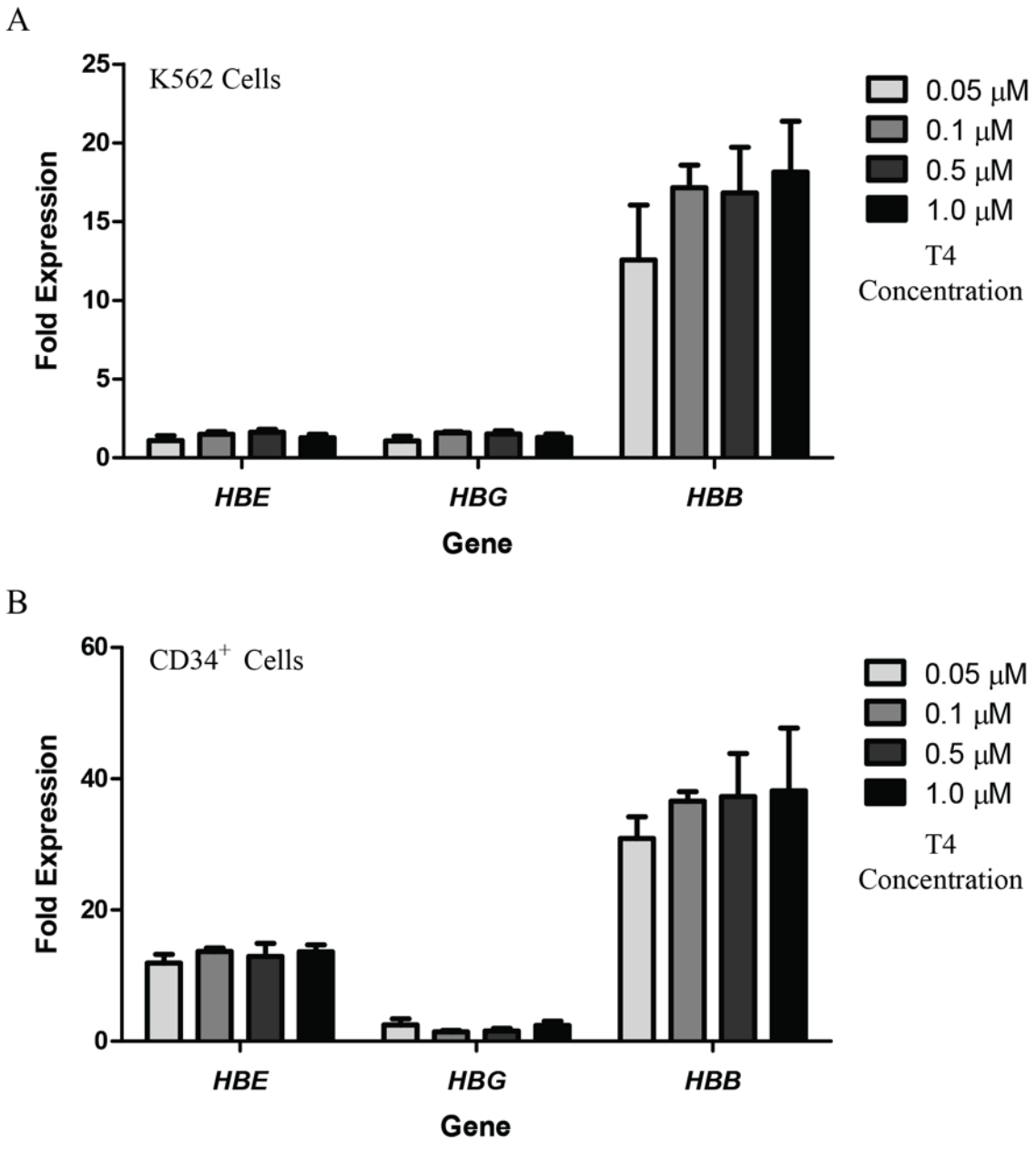
Thyroid hormone receptor pathway members bind directly to globin loci

The thyroid hormone receptor, its heterodimeric binding partner RXR, and its coregulators modulate gene expression through binding to DNA. Therefore, we assessed their binding patterns in K562 cells. In the α-globin locus, all three factors tested bind to the critical α-LCR DNase I hypersensitive site MCS-R3 while both Ncoa1 and RXRα bind to MCS-R1

Figure 3-7. T4 influences *globin* expression in mammalian cells.

Expression of globin genes in T4 treated (A) K562 and (B) differentiated human fetal $CD34^+$ cells. The K562 cells were treated for one day and the $CD34^+$ cells were treated for 3 days. Four concentrations of T4 were used and are indicated by the bar shade. The samples were normalized to vehicle treated controls.

Figure 3-7. (Continued)



(Figure 3-8A). In the β-globin locus, all three factors are bound to the critical LCR region HS 2. HS 4 is also bound by RXRα and TRβ, and 3' HS1 is bound by Ncoa1 and TRβ (Figure 3-8B). The co-binding of these complexes to key conserved regulatory elements strongly indicates they may be functioning by acting directly on the globin locus.

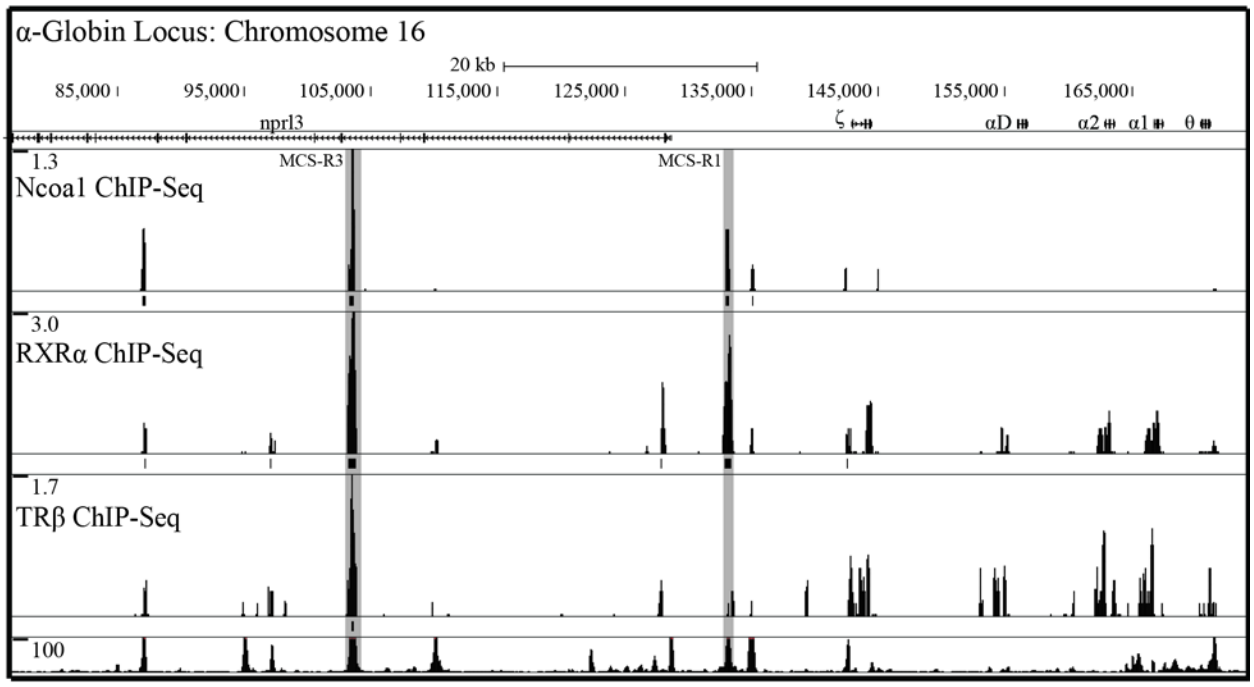
ChIP-seq analysis was also performed on zebrafish larval and adult peripheral blood. H3K27ac mark analysis was performed in cells from both states while RXRα ChIP-seq was performed only in adult cells. In the major locus, regions of differential marking between the adult and larval samples were observed, with higher levels of H3K27ac associated with the region being actively transcribed in that cell type; higher levels on the adult *globins* in adult red cells and higher levels on the embryonic/larval *globins* in the larval red cells (Figure 3-8C). The preferential histone marked region in adult cells encompasses both copies of the two major adult *globins*, α_{a1} and β_{a1}, while the region enriched in larval cells covers all the embryonic/larval *globin* genes present on the locus except one copy of β_{e1}. The minor adult *globins* α_{a2} and β_{a2} are not located in either region of preferential H3K27ac marking. RXRα binding was observed in adult red cells only within the adult specific region of differential H3K27ac. Two distinct peaks are present, centered approximately between the transcriptional start sites of the two α_{a1}/β_{a1} pairs, potentially indicating they are functioning by binding to the bidirectional promoters (Figure 3-8C). At the minor locus, which only encodes embryonic and embryonic/larval genes, only a region where H3K27ac was preferentially bound in larval red cells was observed (Figure 3-8D). The regions encompasses the larval *globins* α_{e5} and β_{e2}, which have similar expression patterns [169]. Two distinct regions of RXRα binding were also observed in the minor locus. One is located within the differentially marked H3K27ac region and between the transcriptional start sites of α_{e5} and β_{e2}, indicating again that RXRα may be functioning through binding to the

Figure 3-8. Thyroid hormone receptor and associated factors bind directly to critical *globin* regulatory elements.

Analysis of the binding of NCOA1, RXRα, and TRβ to the (A) α-globin and (B) β-globin loci in K562 cells. DNase I hypersensitivity from the ENCODE Project is also shown in the bottom track [235]. The bars below each track indicates significant peaks where p<1e-07. Shaded gray regions, and associated labels, indicate known regulatory regions. All tracks are mapped to hg18 and visualized using the UCSC genome browser (http://genome.ucsc.edu). Analysis of the binding of RXRα in zebrafish adult peripheral blood and H3K27ac marks in adult and larval peripheral blood on the (C) major *globin* and (D) minor *globin* locus. The bars below each track indicates significant peaks where p<1e-09. Differential H3K27ac marking between adult and larval peripheral blood was determined and regions preferentially bound in adults (red) and larval (blue) are shown. In some cases annotated genes were renamed in order to adhere to the naming convention, and in cases where a *globin* gene was not annotated, the UCSC BLAT tool was used to locate the ORFs included in the figure. All tracks are mapped to Zv9 and visualized using the UCSC genome browser.

Figure 3-8. (Continued)

A



B

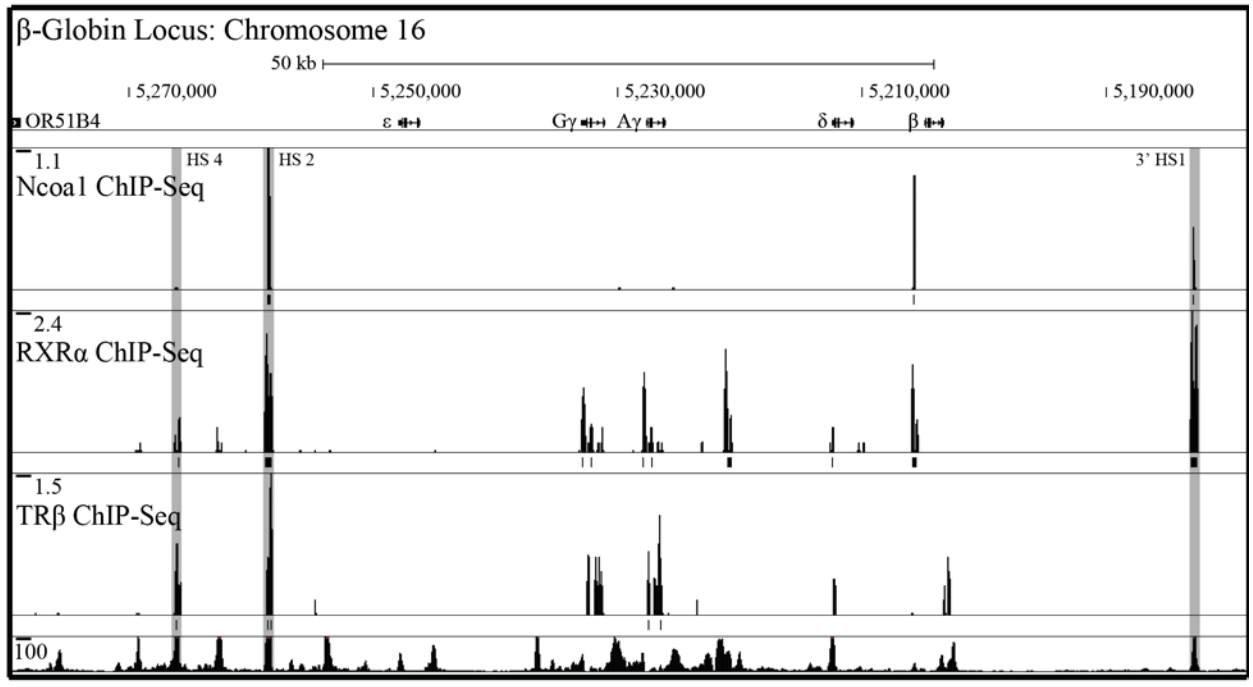


Figure 3-8. (Continued)

C

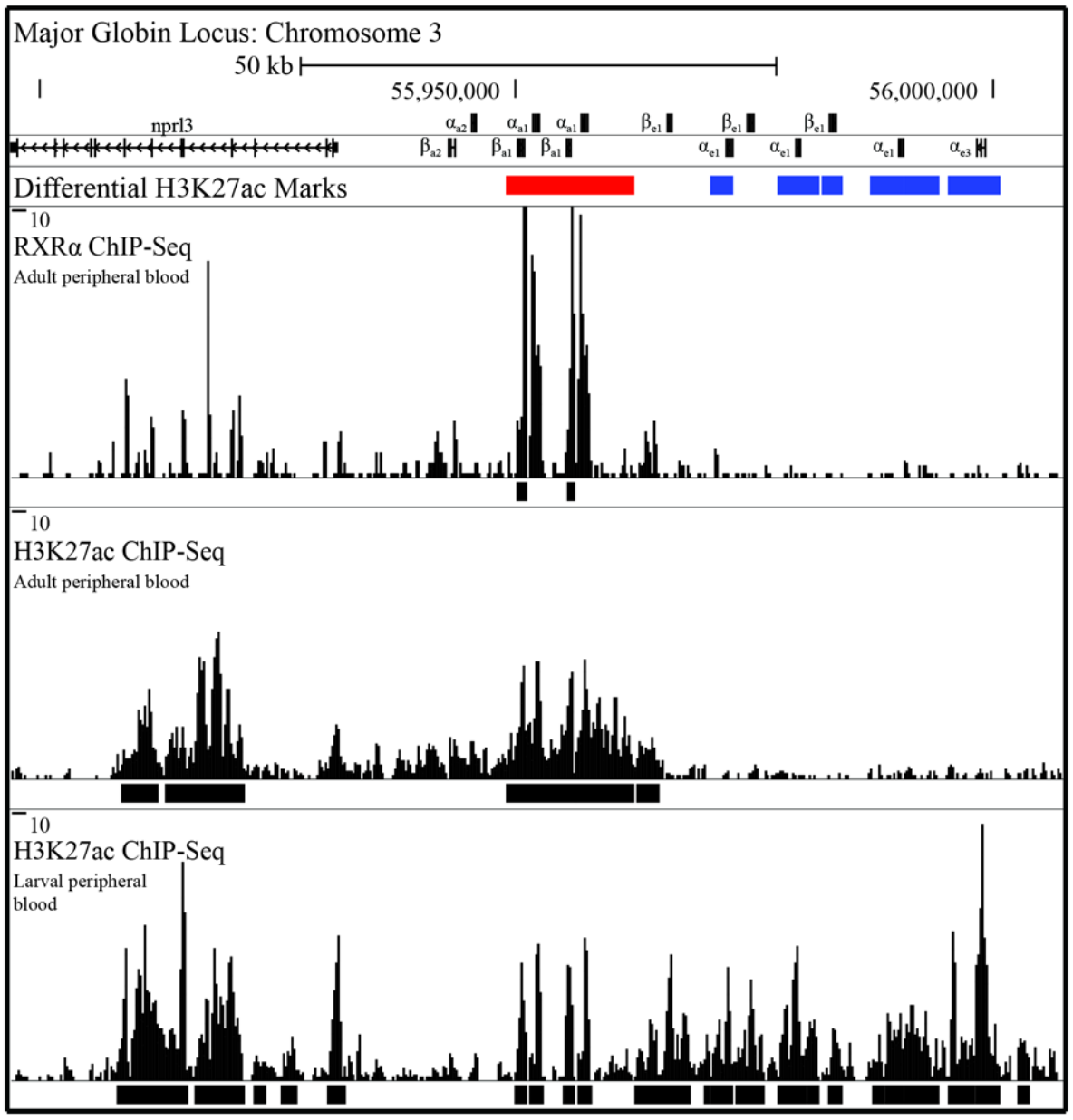
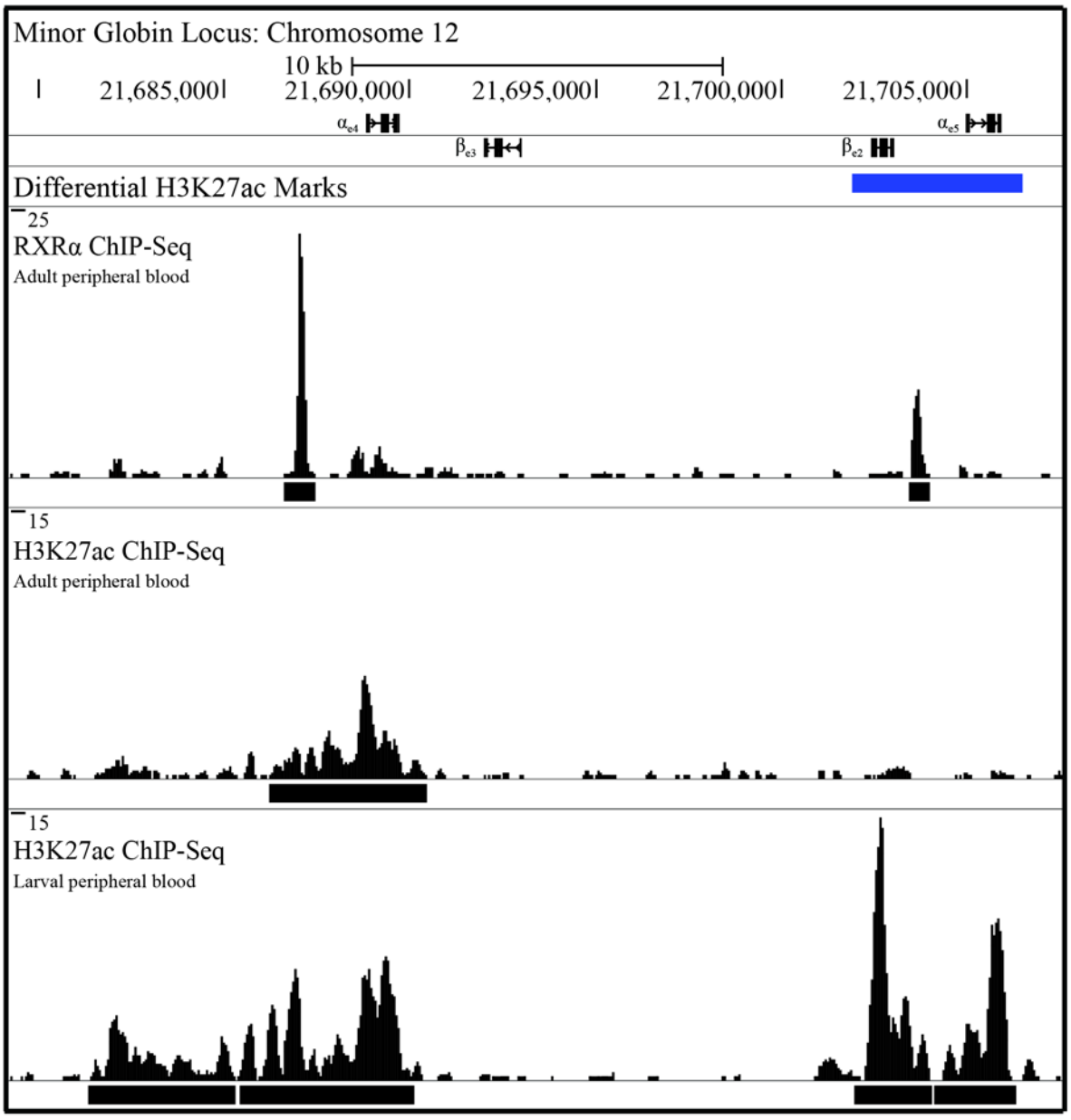


Figure 3-8. (Continued)

D



bidirectional promoters of the *globin* genes. The other RXRα peak is located upstream of α_{e4}, in the region previously identified as the putative minor locus LCR (Figure 3-8D) [169]. These results indicate that critical hormone co-receptor RXRα, and that RXRα binding is associated with critical *cis*-regulatory regions. Therefore the thyroid hormone receptor pathway may be acting directly at the *globin* loci to regulate *globin* switching.

Discussion

The study of hemoglobin switching has been critical to our understanding of general gene regulation. Here we describe two screens, one knockdown and one chemical, that identify two pathways involved in *globin* gene regulation. Down regulation of the Wnt pathway, with the use of Wnt inhibitors or by knockdown of the transcription factor *tcf7l2*, increased the expression of adult *globin* genes. Activation of the thyroid hormone receptor pathway also promoted adult *globin* gene expression in the zebrafish embryo and in mammalian cell culture systems. Conversely, inhibition in the adult zebrafish system induced an embryonic/larval *globin* expression pattern. Changes in globin protein levels were also induced by modulation of this pathway.

The role of Wnt signaling in hematopoiesis has been extensively studied [9,140,142–144,240], but its regulation of *globin* gene switching has not. Here we demonstrate that the pathway does have a role in the erythroid process of *globin* gene expression. While we demonstrate the pathway suppresses adult *globin* expression, further investigation is needed to fully determine its effects on *globin* gene switching and expression.

Thyroid hormone was the strongest single agent inducer of adult *globin* gene expression in the embryo investigated in this study, and adult *globin* was also the most induced gene by

treatment (Figure 3-2, 3-4D). The thyroid hormone receptor pathway is known to be involved in general development, and even the specific changes in globin gene expression that accompany it, since early work in frogs [91,157,241], with later work showing a direct role for the pathway [158]. The hormone is also known to regulate adult *globin* expression in the zebrafish embryo [242], but does not regulate *globin* gene switching in all organisms [243]. The *globin* regulating effect of the thyroid pathway had not been fully characterized nor its interaction with the Wnt pathway identified. Importantly, implications of this effect were also not isolated from general metamorphic changes, extrapolated to the adult state, or translated to mammalian systems.

Here, we demonstrate the ability of the thyroid hormone pathway to regulate both *globin* gene and protein expression outside of the context of metamorphosis. By RNA-Seq analysis, a very limited set of genes is altered by T4 treatment (Figure 3-4D), indicating that, at the time points and/or doses used, full metamorphosis is not taking place. The role of the thyroid hormone pathway in the maintenance of the adult *globin* state in addition to the establishment of it during metamorphosis, as demonstrated by the thyroid hormone receptor antagonist inducing embryonic/larval expression, also speaks to the pathway's role beyond metamorphosis. This is consistent with observed T3 levels in the zebrafish peaking during metamorphosis, but also remaining at a higher concentration in the adult than the early embryo [156].

T4 treatment induced adult *globin* expression with expected kinetics. With embryo treatment, mRNA is detectable at least as early as 24 hpf, but adult protein is not yet detectable. 24 hours later, at 48 hpf, induced adult globin protein is clearly detectable (Figure 3-5A). This protein induction, as well as the induction seen at the larval stage and in the adult stage (Figure 3-5B, C, 3-6C, D), is critical to demonstrating that no additional layers of regulation restrict the effect of the thyroid hormone pathway after mRNA induction. While we do not observe it

directly, we are able to infer that chain balance at the protein level, critical for producing functional hemoglobin, is not disrupted as we see the *α-* and *β-globin* genes being coordinately affected at the mRNA level (Figure 3-1H, 3-4A, B, 3-6A, B). These results show that the thyroid hormone receptor pathway could potentially be modulated clinically without disrupting normal hemoglobin formation and the oxygen carrying function of red cells, which would be important for therapeutic applications.

The co-expression of T4 induced adult *globin* expression potentially indicates a model where the *globin* genes are switching within an individual red cell. Because we observe cells that are expressing both the normal *globin* and the induced *globin*, as well as cells expressing only the induced *globin* (Figure 3-5D), it is possible that the co-expressing state represents a transitional state. This transition would be preceded by the single expression of the normal *globin* and followed by expression of the induced *globin*, all within the same cells. This process would be akin to "maturational" *globin* switching rather than the cell replacement model typically seen in the normal larval to adult switching process [169]. Similarly, the presence of coexpressing normal adult *globin* and thyroid hormone pathway modulation induced embryonic/larval *globin* cells alongside side non-coexpressing cells of each type (Figure 3-6E) indicates an analogous process is likely occurring in the Dron treated adults. In addition to also supporting the idea of a "within cell" transition, it also demonstrates that the adult state is not "locked" in. While normal, endogenous processes, most likely involving the thyroid hormone pathway, allow the progress from embryonic, to larval, to adult *globin*, under normal conditions, a transition from the adult state to an embryonic or larval does not occur. This would allow for the possibility that a "locked" cells state, epigenetic or other mechanisms, could exist that would not allow reversal of the process. As we are able to achieve this reversal, it demonstrates this

condition is not present and the model is malleable and ideal for studying this abnormal process. These transitions of *globin* genes within a single cell would be consistent with the model that the thyroid hormone receptor is binding to the *globin* locus and upstream enhancers to directly affect *globin* gene expression. Utilizing T4 to induce adult globin expression in fetal-like cells is also advantageous to those attempting to generate mature, adult globin expressing erythrocytes from iPS or ES cells, which using current protocols often results in cells expressing embryonic or fetal *globin* genes [244].

The identification and study of these pathways involved in hemoglobin switching using the zebrafish system also demonstrated the utility of the zebrafish as a model system for the study of this process. The benefits of screening in the system are well documented [109,110,112,187], but the ability to translate findings to mammalian systems in the field of hemoglobin switching were not yet established. Here, we demonstrate the conservation of these pathways from the mammalian system to the zebrafish embryo and adult systems. This allows the study of this complex process in the context of a whole organism model more similar to, and malleable than, other whole organism (i.e. mice) models [169].

Clinical implications also exist for the use of thyroid hormone receptor antagonists in the treatment of hemoglobinopathies, such as sickle cell anemia. Sickle cell disease is caused by a mutation in the adult β-globin gene that promotes polymerization of the translated protein, which in turn causes the distinctive sickling of the red cells and downstream disease symptoms. Increased levels of fetal hemoglobin ameliorate the symptoms of sickle cell disease by impeding polymerization, making pharmacological induction of fetal hemoglobin a logical mechanism to pursue for treatment of this disease [79,176,211]. In the zebrafish, we are able to induced fetal hemoglobin in the adult state with the thyroid hormone receptor antagonist dronedarone, a

United States Food and Drug Administration approved molecule for heart arrhythmia [245]. While we were only able to achieve a modest increase in embryonic/larval hemoglobin in peripheral blood (Figure 3-6C), we did observe a larger numbers of positives cells in the kidney marrow (Figure 3-6D). In addition, large fold increases of embryonic and larval *globin* mRNAs were observed in the kidney marrow (Figure 3-6B). This may indicate that peripheral blood levels may rise further as these cells begin to enter circulation. The short duration of the experiments (5 days) in conjunction with the relatively long half-life of red cells, approximately 120 days in humans [246] and up to 270 days in carp [247], may mean treatment over a longer period, which would allow for additional red cell turnover, could increase the levels observed in peripheral blood, and make thyroid hormone receptor antagonists a viable potential treatments to pursue for sickle cell disease. There is also limited evidence that, despite the feedback loops and other mechanisms regulating the human thyroid hormone axis, non-anemic hyper- and hypothyroid patients have abnormal *globin* expression profiles [248]. Continued study of these areas may open additional avenues for clinical investigation into pharmacological means to increase fetal hemoglobin by modulation of the Wnt and thyroid hormone receptor pathways.

Chapter 4

Discussion and future directions

Hemoglobin switching is a complex and tightly regulated developmental process. Mutations in the coding and non-coding, regulatory regions of the hemoglobin loci in humans directly lead to diseases ranging from mild to life-threatening. The studies completed in this thesis aim to better understand the complex molecular coordination required for normal hemoglobin expression, switching, and overall function. The zebrafish was characterized as a model for hemoglobin switching by annotating the *globin* loci, determining the full complement of *globin* genes expressed, and the developmental timing of the switches between these *globins*. The key regulatory element, the locus control region (LCR), for the major locus was also identified, allowing the generation of a red cell specific transgenics with a GFP reporter. With the established standard for *globin* expression, it was then possible to screen for regulators of *globin* switching. Both morpholino and chemical screens were undertaken, and the Wnt and thyroid hormone receptor pathways were identified as regulators of *globin* expression. Manipulation of these pathways can modulate globin protein expression in both the embryo and adult zebrafish. Studies in two cell culture systems, differentiated human primary fetal $CD34^+$ cells and the erythroid leukemia line K562, demonstrated that these pathways also regulate globin gene expression in mammals. Through mechanistic studies carried out in both the zebrafish and human systems, the thyroid hormone receptor pathway is likely acting directly on the *globin* locus to influence gene expression. This work has implications for furthering a better general understanding of hemoglobin switching and gene regulation, and, through this improved understanding, lay the foundation for better, more targeted treatments for some hemoglobinopathies.

Modeling Hemoglobin Switching in the Zebrafish

Hemoglobin and hemoglobin switching have been studied since their first descriptions in 1840 and 1934 respectively [20,59]. Significant progress in understand these topics has been made in the nearly 175 interceding years, many in conjunction with new models, and new techniques that can be used to study them. The most common models currently used to investigate these topics vary from immortalized cells lines, primary cells from model organisms and humans, traditional model organisms, and transgenic systems.

Human and murine cell lines such as the human erythroid leukemia line K562 and the murine erythroid leukemia line GM979 have been used to study this developmental process, but are often limited as they do not directly reflect the normal state of any naturally occurring cell state, nor interact in a proper cellular context. The ability to obtain primary cells, such as human $CD34^+$ cells, has allowed some *in vitro* studies to more directly investigate a normal cellular state. However, once in culture, these cells also lack the normal interactions of their endogenous environment, and face technical limitations associate with primary cell cultures. Whole model organisms, such as mice, allow these processes to be investigated in the correct context, although are not directly comparable to human systems. Most other non-primate species experience only one *globin* switch during development. The fetal stage expression of γ-globin evolved more recently during primate evolution [98,249]. In the murine system, this distinction has been investigated and found, at least in part, to be driven by differential regulation and expression of the key globin regulatory protein BCL11A [64]. Practical and technical considerations also limit the use of these organisms for screening and studies of early developmental steps. Transgenic systems, particularly the β-globin locus transgenic mouse [52], have been heavily, and successfully, used to study hemoglobin switching, but face many of the same practical and

technical limitations as non-transgenic models. They also do not exactly recapitulate the human hemoglobin switching regulatory system as the murine environment results in the aberrant expression of the human genes [171].

Characterizing the zebrafish as a model system

The zebrafish system combines a number of advantages of whole organism, transgenic, and cell cultures systems. The ability to quickly and efficiently screen by knockdown, chemically, or genetically in the context of a complete developing organism is ideal for investigating a complex developmental process. However, in order to best utilize the system for the study of hemoglobin switching, the model needed to be better characterized and further developed.

While hematopoiesis in the zebrafish, and even *globin* gene expression, had been studied in the fish, a detailed analysis of the genetic, epigenetic, temporal, and regulatory aspects had not been completed. By assembling and annotating both the major and minor *globin* loci from BACs and other genomic assemblies, the complete set of *globin* genes present in the zebrafish was able to be identified. In addition, many of proximal regulatory elements were identified as shared, bi-directional promoters paired with an *α*- and a *β-globin*, suggesting a simple mechanism to maintain the proper, critical balance between *α*- and a *β*-chains. While earlier work had demonstrated the presence of two *globin* loci in the zebrafish [124] and their arrangement in a "head-to-head" fashion separated by a bi-directional promoter [30,32], these features had only been identified for isolated pairs of *globin* genes, and did not map and annotate the structure of both loci. The genomic and regulatory organization observed is distinct from that seen in mammal. The mammalian system has two separate *globin* loci, genetically segregating the α-

and β-like genes [38]. This arrangement implies an additional level of coordination between the loci to maintain chain balance that is unlikely to be present in the zebrafish system.

To establish the temporal component of the hemoglobin switching process, a combination of *in situ* hybridization (ISH) and real-time quantitative PCR (RT-PCR) were used. These methods corroborated each other to establish the normal, base line *globin* expression pattern. Establishing this pattern was essential in order to be able to evaluate perturbations. The switching process begins with the onset of *globin* expression, switches twice, and completes around 32 dpf. These results are roughly consistent with earlier work completed to determine the timeline of *globin* expression in the zebrafish including the stage of the onset of *globin* expression and the timing and location of the expression of different *globin* genes though 3 dpf [30]. However, in the adult stage, the results observed here differ from those in previous work. Through the analysis of cDNA libraries and HPLC for globin proteins, five major adult globin proteins were identified: two β-globin and three α-globin [32]. While this work also identifies two adult *β-globins*, the results would suggest a substantially larger percent of the total adult *β-globin* to be composed of β_{a1} rather than β_{a2} based on the ratios observed in the mRNA expression. While the expression ratio similarly favors β_{a1} at the protein level, β_{a2} contributes a higher percent of the total protein [32]. This could represent another layer of *globin* gene regulation, differential translation of mRNA species, or natural variation in techniques used. The more striking difference observed is in *α-globin* expression. Only a single *α-globin* species was found to contribute to the adult red cell population in this work compared to the three observed previously [32]. This discrepancy could be the result of posttranslational modifications altering the protein, the limitation of RT-PCR to detect all possible variations in cDNAs, or technical caveats leading to preparation related protein degradation. Another potential explanation for the

observed variations could be presence of multiple haplotypes in lab strains of zebrafish. Unlike laboratory mice strains, the zebrafish lines are not inbred and efforts were made to maintain genetic diversity within stock populations, so haplotype variations within a line are possible. During the sequence analysis of PAC and BAC clones that were made from genomic DNA samples isolated from the same zebrafish line, SNPs and other variations were observed in the major *globin* locus. Almost all coding sequence variations were synonymous, but as we do not yet fully understand where all of the *globin* gene regulatory elements are and their functions, it is possible that non-coding mutations could have substantial differential effects on *globin* gene expression between individuals and families that inherit different hyplotypes of the *globin* locus in zebrafish stock populations (Yi Zhou, personal communication). While this work clearly maps the *globin* loci and expression patterns of the *globin* gene mRNAs, further analysis at the protein level is needed to address this variation.

Taken together, these results demonstrate the presence of two switches and three stages of *globin* expression analogues to those observed in humans, but the human switching process is completed between 10 and 11 months of development, substantial later than in zebrafish. While in mice the process is completed between 13 and 14 days, only one switch occurs in the murine system [171]. The zebrafish provides an accelerated, two switch system to study the hemoglobin switching process. In many cases changes to *globin* expression are correlated to changes in the location of erythropoiesis and the wave, primitive or definitive, of hematopoiesis [4,6,55]. This work, and previous analyses [30,32], generally support that model as the timing of the major *globin* switches coincide with the changes in location of erythropoiesis from the ICM, to the CHT, and finally to the adult kidney [27]. However, other variations in *globin* expression, particularly the shift in expression away from β_{e3}, support a role for maturational *globin*

switching [26]. Therefore, it is likely that both cellular turnover and maturational *globin* switching contribute to the overall switching pattern observed during the development of the zebrafish.

Understanding globin regulation in the zebrafish

In addition to proximal promoters, *globin* genes are regulated by long-range enhancers in mammals, known as locus control regions (LCR). These regions are characterized by critical DNase I hypersensitive sites and the binding of critical red cells transcription factors (e.g. GATA1). In order to locate, characterize, and utilize an analogous region in the zebrafish, DNase I hypersensitivity and Gata1 ChIP-seq were performed. A putative major locus LCR was determine bioinformatically and tested experimentally. As the region was able to confer specific and robust reporter expression, it was characterized as the zebrafish major locus LCR. However, the region does not contain sufficient information to result in proper, or any, switching behavior, even when combined with a native, temporally regulated zebrafish *globin* promoter. This indicates the *cis*-regulatory elements sufficient for correct temporal control have not been completely elucidated. In the mammalian system, the LCR, driving expression of a *globin* gene, is also sufficient to result in high levels of specific expression [35,37]. However, as observed here, normal developmental/temporal regulation is disrupted without the complete local organization found in the native locus [37,49,50,52]. The LCR discovered in the zebrafish is an α-LCR. The major locus is syntenic with the human α-globin region, like other fish such as the puffer fish [34], and contains the conserved gene *nprl3* (C16orf35) immediately upstream of the *globin* gene cluster. As with the human α-LCR [42,99], the zebrafish major locus LCR is located within the introns of this gene. This is consistent with the known evolution of the loci; the zebrafish loci diverged from mammalian precursor *globin* loci, before the divergence of the α

and β clusters [33]. It is also consistent with the observed homology between the LCRs of other fish and humans [42].

Screening in the zebrafish model

With the establishment of the *globin* loci, the normal *globin* expression pattern, and the zebrafish major locus LCR, it was possible to take advantage of one of the most significant advantages of the zebrafish system: the ability to perform large scale screens on a whole organism. The ability to generate a large number of optically clear, externally fertilized embryos quickly, and then easily assay temporal and spatial gene expression by *in situ* hybridization allows both chemical and knockdown screens to be performed in a high-throughput manner [109–112]. It is also possible, while in lower throughput manner, to utilize the adult zebrafish for analysis [113].

Through accentuating the advantages and mitigating the disadvantages, and through the development of additional tools such as the LCR-GFP transgenic line and the zebrafish adult globin specific antibody, morpholino and chemical screens have been successfully completed. The success of these efforts demonstrates that the zebrafish, both the embryonic and adult systems, are appropriate systems for the productive study of hemoglobin switching.

Thyroid Hormone and Wnt as Regulators of Hemoglobin Switching

Identification of novel regulators of hemoglobin switching

Once the zebrafish was established as a screening system for hemoglobin switching, two screens were designed to identify genes and pathways in its regulation. A knockdown screen was designed using morpholinos injected at the one cell stage, with the embryos being assayed at

24 hpf. This time point was selected to allow the maximum amount of time for the morpholinos to affect to *globin* gene expression prior to circulation beginning. This would allow the easiest scoring of the embryos as all the red cells are located in the intermediate cell mass [27]. An increases in the adult *globin* gene α_{a1} was selected as the readout for the screen because: (1) the adult and larval *globins* have a low background level of expression at 24 hpf and small increases over a minimal base line are easier and more reliable to score compared to small increases or decreases in a highly expressed genes, such as the embryonic and embryonic/larval *globins*, (2) normal adult expression does not being until at least 17 dpf, leaving a large window to accommodate staging errors if they were to occur and would imply any observed change indicated an important alteration to *globin* regulation rather than a minor acceleration, and (3) α_{a1} has a robust, reliable *in situ* probe for the detection of abnormal expression. However, this approach does have limitations. While the number of α_{a1} positive cells was able to be sensitively detected, the absolute changes in α_{a1} expression levels were not assessed, nor were the changes to any other *globin* genes. This increased sensitively over RT-PCR, which would be used to detect changes in all *globin*, was most likely critical to detecting the changes observed in the knockdown screen as the initial phenotypes observed were subtle. While chemical screening is extremely high-throughput, there are practical limitations for morpholino based knockdown screens. In order to generate a candidate gene list enriched for potential regulators of *globin* switching, data from murine microarrays on erythroid cells were utilized. The hypothesis was that transcription factors regulating *globin* changes would be differentially expressed in cells expressing different *globins*. Furthermore, as factors would be knocked down in embryos and adult *globin* would be assessed, factors were identified that would more likely generate this phenotype; factors that were over expressed in the fetal, adult, or fetal and adult stages compared

to the embryonic state were selected. Mouse cells and microarrays were utilized as sufficient markers are not available to separate all of the desired cells types in the zebrafish system. This introduced additional potential for cross-species difference to affect the validity of the screen.

After generation of the candidate gene list, completion of the knockdown screen, and completion of the chemical screen, four genes and number of chemicals had been identified. In order to determine which pathways to investigate further, the results of both efforts were compared. The largest increase observed in the knockdown screen was the combination of *tcf7l2* and *ncoa1* knockdown. As these are associated with the Wnt and nuclear hormone receptor pathways respectively, and two significant results from the chemical treatments were XAV939 and T4, representing the Wnt and thyroid hormone receptor pathway, a member of the nuclear hormone receptor superfamily [146], respectively, these two pathways were pursued. Specifically, T4 treatment was the most substantial single agent inducer of adult *globin*, so was further selected for additional testing.

Monoclonal antibody development

While the zebrafish model was well characterized during the course of this work, tools to assess protein expression were still limited. HPLC and mass spectroscopy have been utilized to assess zebrafish *globin* expression at the protein level [30,32], but are limited for assessing protein expression and co-expression at the cellular and whole-mount level. An antibody against β_{e1}, an embryonic/larval globin, is commercially available (Anti-Hbbe-1.1 (NT), Z-Fish; Catalog #55608, Anaspec, Fremont, CA), but no zebrafish adult globin specific antibody is available. An adult globin specific antibody was developed in conjunction with the Dana-Farber Cancer Institute Monoclonal Antibody Core (Boston, MA) by immunizing mice with purified adult

zebrafish red cell lysate. As peripheral blood is predominantly red cells and globin is the predominant protein produced by red cells, peripheral blood was collected from adult fish, lysed, and membrane fragments removed by ultracentrifugation. After multiple rounds of boosting, spleen cells were collected and fused to form hybridomas. Clones were screened using a combination of western blotting and blood smear antibody staining to select a clone specific to adult globin. Whole protein rather than a peptide fragment was used for immunization in order to increase the likelihood that the antibody would work on protein in the native confirmation, such as in blood smears and whole-mount staining. With the development of this additional tool, it facilitates future globin expression studies in the zebrafish as embryonic and adult globins can now be detected and distinguished at the mRNA and protein levels.

Role of Wnt in adult globin expression

Inhibition of the Wnt pathway, with the inhibitor XAV939 or by knocking down *tcf7l2*, in the embryo leads to an increase in adult *globin* expression. These results indicate that Wnt is involved in the maintenance of the embryonic state, or potentially that the absence of a Wnt signal in the blood at this time triggers the activation of another signaling pathway that promotes the adult *globin* state. There is no known role of *tcf7l2* or Wnt directly in hemoglobin switching, but the pathway is integrally involved in hematopoiesis [140,143–145], increasing the likelihood that such a role would exist.

Role of thyroid hormone in embryo-to-adult globin switching

Activation of the thyroid hormone receptor pathway with T4 treatment increases adult *globin* expression in the embryo. This finding is consistent with the literature on the role of T4 in the induction of metamorphosis [155,215], and even the direct role of hemoglobin switching

in amphibians [250]. The direct role observed in amphibians was demonstrated by the addition of T3 to tadpoles followed by assessment of the expression of both tadpole (embryonic/larval) and frog (adult) globin genes. The expression of frog globin is induced by T3, but in never in the same cells as tadpole globin, similar to the coexistence of distinct tadpole globin expressing and frog globin expressing cells observed during natural metamorphosis. This, and additional experiments, indicate that in the amphibian system T3 induces frog globin expression in progenitor cells that undergo DNA replication prior to the onset of globin expression [250]. The upregulation of adult *globin* with T4 treatment in zebrafish has also been reported as part of a microarray study on the effects of bisphenol A in the environment [242], but the implications for hemoglobin switching have not been investigated. The ability to increase the amount of both adult *globin* mRNA and protein in embryos and larva is critical, as it demonstrates that the effects of the thyroid hormone pathway is not limited in inducing adult *globin* that can be incorporated into functional adult hemoglobin. The effects of the pathway would be less interesting if it induced only mRNA and not protein. The induction of mRNA, as observed by *in situ* hybridization, occurs in a short period of time, approximately 19 hours between the initiation of T4 treatment at 50% epiboly and the assay at 24 hpf. Induced adult globin protein is not detectable at this stage (24 hpf), but is detectable by 48 hpf. When treating larva, induced adult globin positive cells were detectable 24 hours after the initiation of T4 treatment (15 dpf-16 dpf). Together these data indicate that T4 is able to induce adult *globin* mRNA in under 24 hours, with protein beginning to appear in approximately 24 hours. When this induced adult globin protein is detected, it can be detected in some cells that are also expressing the embryonic/larval globin β_{e1}. One potential explanation for this observation is that cells already expressing embryonic/larval globins are being induced to express adult globin and transitioning thought the

co-expressing state observed. This observation is in contrast that that observed in amphibians, where thyroid hormone induced frog globin is not co-expressed in the same cell as tadpole globin [250]. This would imply that the cell type that is affected is already expressing embryonic globin and not an early progenitor, but further study is needed to determine the target cell type(s) as well as the true nature of the co-expressing cells. The interactions between the two pathways identified that induce adult *globin* expression in the embryo, the thyroid hormone receptor and Wnt pathways, have been characterized [165–168], but not in relation to hemoglobin switching. Further study is needed to better characterize this relationship in this context to develop a more complete picture of hemoglobin switching regulation

Role of thyroid hormone in adult-to-embryonic/larval globin switching

Inhibition of the thyroid hormone receptor pathway with dronedarone treatment increases embryonic and larval *globin* expression in adult peripheral blood and kidney marrow. This result implies that the thyroid hormone receptor pathway is involved in the maintenance of the adult *globin* state in addition to its establishment. Similar to the co-expression of embryonic/larval and induced adult globin in the embryo and larval, a number of adult cells co-express adult globin and induced embryonic/larval globin, β_{e1}. This many imply the existence of a transitional state and implicate a globin expressing cell type as the target of globin modulating thyroid hormone receptor signaling. This ability to modulate thyroid hormone receptor signaling to induce embryonic and/or larval globin is particularly interesting as it may have implications for the clinical treatment of some hemoglobinopathies [176].

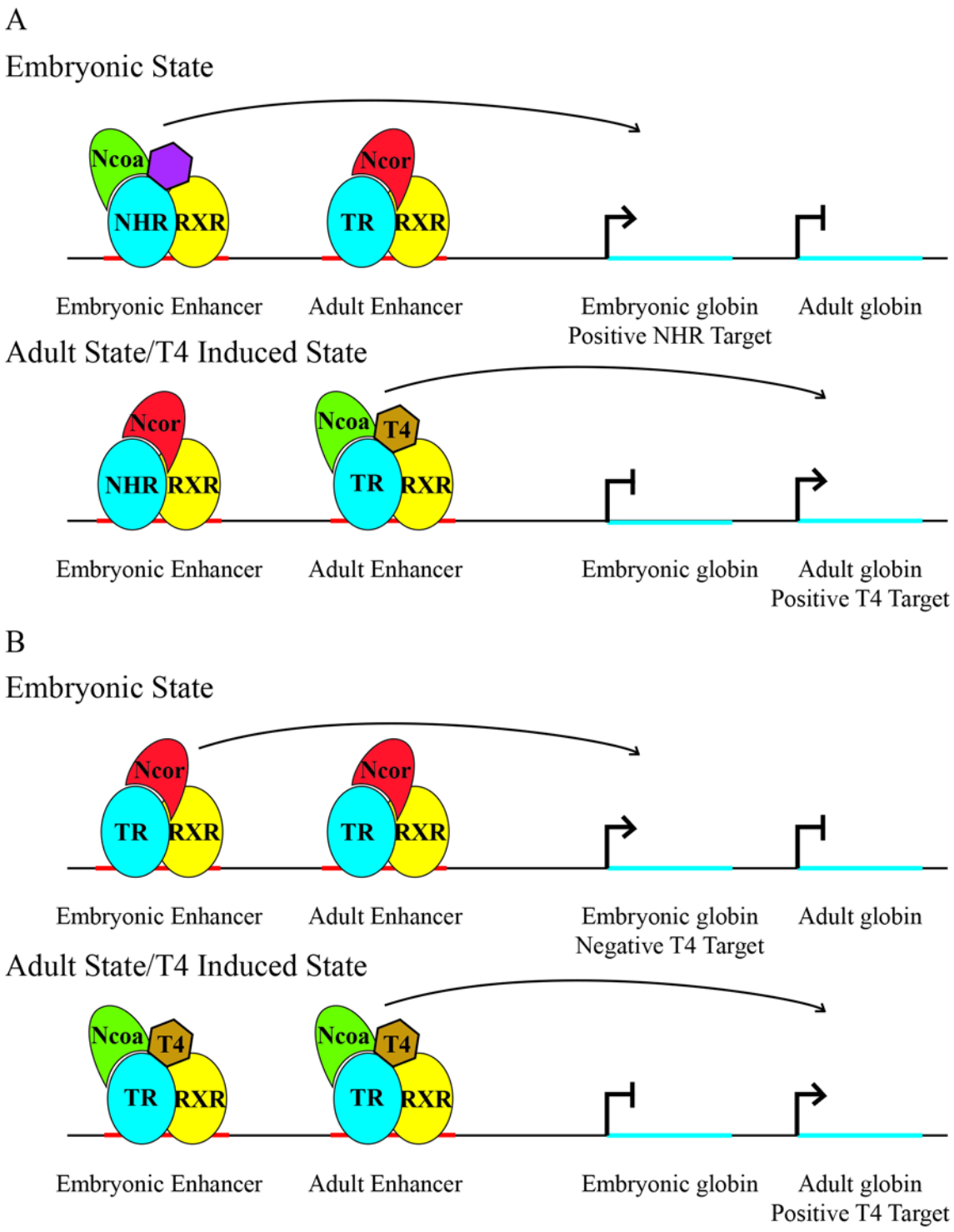
Role of thyroid hormone in mammalian globin switching

T4 treatment of fetal globin expressing K562 cells and differentiated primary human fetal $CD34^+$ cells increases adult β-globin expression. While it has been speculated that there is conservation of a metamorphosis-like process in amniotes, including mammals [159], the thyroid hormone receptor pathway has not been observed to regulate globin switching. This link, while still currently limited to cell culture experiments, is critical as it demonstrates that the observations made in zebrafish, a system that does undergo traditional metamorphosis, are still valid in a non-metamorphosing mammalian species. This not only provides support for the continuation of this line of work, but also further justifies the zebrafish system for the study of hemoglobin switching. The evidence these systems provide to show that the thyroid hormone receptor pathway may act directly on the globin locus is also critical. While the thyroid hormone receptor pathway can have numerous secondary effects in various tissue types, a direct, ligand induced effect mediated though a receptor directly bound to a target gene would be consistent with known thyroid hormone receptor functions in other tissues and systems [150]. As the regulation of globin switching is multilayered and complex, identifying a pathway that may be direct could: (1) allow for an increased effect as fewer other pathways can compensate and (2) limit off target effects, at least within the target cell type. While the data presented here still allow for a few potential models of this direct interaction, one model would restrict the role of T4 bound to TR an NCOA factor to inducing adult globin expression and invoke another, yet unidentified, nuclear hormone receptor pathway to promote embryonic globin expression in the embryo (Figure 4-1A). Another model would be that TR directly regulates both embryonic and adult globin expression, with the embryonic *globins* being a negative target and the adult *globins* being appositive target (Figure 4-1B). However, the thyroid hormone axis in humans is itself

Figure 4-1. Models of thyroid hormone receptor action on *globin* loci.

(A) In the embryonic state, embryonic globin genes are activated by a yet unidentified nuclear hormone receptor complex containing an Ncoa factor, and adult gene transcription is repressed/not activated by a TR complex containing an Ncor factor. In the adult state, T4 is present and the ligand for the unidentified nuclear hormone receptor is not. Gene transcription of the adult globin genes is activated by a TR complex containing an Ncoa factor, and embryonic globin transcription is repressed/not activated by a complex containing an Ncor factor. (B) In the embryonic state, embryonic globin genes are activated by a TR complex containing an Ncor factors, and adult globin gene transcription is repressed/not activated by a TR complex containing and Ncor factor. In the adult state, T4 is present and gene transcription of the adult globin genes is activated by a TR complex containing Ncoa, and the embryonic globins genes are repressed/not activated by a TR complex containing an Ncoa factors. (Ncoa – any member of the nuclear hormone receptor coactivator family; Ncor – any member of the nuclear hormone receptor corepressor family; NHR – a yet unidentified nuclear hormone receptor; TR – either thyroid hormone receptor; RXR – any member of the retinoid X receptor family; T4 – thyroid hormone; purple hexagon – ligand for the yet unidentified nuclear hormone receptor)

Figure 4-1. (Continued)



complicated, regulated by feedback loops, and can have major systemic effects, so may not be an ideal therapeutic target.

Clinical implications for hemoglobinopathies

With the evidence that the thyroid hormone pathway both establishes and maintains the adult-like *globin* state in zebrafish, and that the function of this pathway may be conserved in mammalian cells, it possible that the thyroid hormone receptor pathway is also involved in the maintenance of the adult pathway in humans. This would imply that inhibition of the pathway would induce embryonic and/or fetal hemoglobin production in hemoglobinopathy patients, ameliorating some symptoms. Differences in thyroid hormone levels have been correlated to the expression levels of particular hemoglobin genes in humans [248], but further testing is needed to better characterize and quantify these correlations, particularly directly in patient populations. In addition, the number of embryonic globin positive cells induced with dronedarone treatment is low compared to what would be required to achieve a clinical effect [176]. However, the short treatment window, retro-orbital route of administration, cross-species differences and other factors could contribute to a lower level observed in zebrafish than could be obtained in patients. In addition, the molecule, exact target in the pathway, or treatment cocktail could be optimized for use in humans. It is unlikely manipulation of the global thyroid system would be a safe, effective method of eliciting such a specific response. This work demonstrates that targeting specific factors in the pathway, such as Ncoa1 or only one of the thyroid hormone receptors, maybe a more targeted way to maximize the desired response while minimizing the off target effects. It may also be possible to specifically target erythroid cells in order to limit systemic effects by treating cells *ex vivo* or utilizing antibody targeting technologies [251]. While the thyroid hormone pathway may not be an ideal target for pharmaceutical treatment of the

hemoglobinopathies at this time, and substantial additional work in model systems, and humans, will be needed, the data presented here support further investigation into its therapeutic potential.

Future of Hemoglobin Switching Research in the Zebrafish

Identification of additional new regulators

With the characterization of the zebrafish as a model for hemoglobin switching complete and the success of the morpholino knockdown and chemical screens, these data demonstrate that the zebrafish is an ideal model for identifying and studying regulators of *globin* gene expression. A natural extension of this work would be the completion of an overexpression screen. The knockdown screen completed in this thesis utilized candidate genes generated from the analysis of microarrays data on mouse red cell samples collected form yolk sac, fetal liver, and adults. These samples express different *globin* genes, and therefore, potentially, differentially express *globin* gene regulators. During the process of identifying candidate genes for the knockdown screen, genes for which a knockdown in the embryo would result in a state more closely resembling that of the adult were selected. The remaining genes, those for which overexpression would result in a state more closely resembling that of the adult, have therefore already been identified. This list is therefore likely to be enriched for regulators of *globin* gene expression as the list examined in this work yielded a high percentage of "hits" (~23%). The list of candidate genes to overexpress also contains known critical regulators of *globin* switching *bcl11a* and *sox6*, which function together to act as suppressors of fetal hemoglobin in mammals [63–65], consistent with overexpression in an embryo leading to increased levels of adult *globin* expression. In addition, preliminary data in the zebrafish shows that overexpress of *bcl11a* does in fact lead to a significant increase in the number of *α-globin* mRNA positive cells in the

embryo at 24 hpf (Janelle Lambert, appendix). While overexpression screens are typically more difficult to establish, execute, and interpret than knockdown screens, they have been completed successfully in zebrafish [113]. In addition to having already generated an enriched candidate gene list, there are also a number of tools now in place to expedite such a screen. First, a new Gateway recombination cloning technology (Life Technologies, Grand Island, NY) destination vector was generated. This vector contains the draculin (drl) promoter, which drives expression in the mesoderm, including red cells (Christian Mosimann, manuscript in submission), driving the expression of GFP as a reporter in addition to the standard recombination sites used to easily clone in a desired promoter/gene of interest combination. This vector with an additional, smaller drl promoter driving another reporter gene has been successfully cloned and expressed in zebrafish (data not shown). Use of this reporter would aid a screener in identifying and selecting successfully injected embryos, which has been critical in similar screen [113]. An improved potential screening assay could also be used in place of *in situ* hybridization. Now that an approximate time line has been established for induced globin protein expression (just over 24 hours), whole mount fluorescent antibody staining using the newly developed adult globin specific antibody could be utilized to assay for changes in protein expression. While *in situ* hybridization was successfully used in this work, whole mount antibody staining is a faster, more robust protocol, and throughput could be increased by utilizing it. Also, the screen would directly read-out genes that result in changes to protein levels, which is critical for follow up investigations. The scoring system used in this work, in addition to the tools developed to assay and quantify the results, could also be adapted to an overexpression screen, reducing the need to completely redevelop these aspects. Similarly, the presence of a positive control, identified in this work, could allow for expedited screen development and more accurate results. Such an

overexpression screen might also have an advantage over a knockdown screen; translation to increasing fetal hemoglobin in human patients may be more straightforward, as pharmaceutical inhibition of targets is typically more common than activation.

New technology developed since the initiation of this work, the CRISPR-Cas system, could also be utilized to screen for additional regulators of hemoglobin switching. The system utilizes a bacterial derived nuclease and guide system to generate genomic mutations and deletions in zebrafish and other systems [252–256]. Once genes were knocked out, the phenotype at each developmental stage could be evaluated, whereas morpholinos are limited to only early stages. In addition, the knockdown would be complete and able to be verified by PCR. However, this could also limit the analysis of essential genes as different doses of the target cannot be achieved.

Further characterization of the role of the Wnt pathway

In addition to identifying new pathways, the role the Wnt pathway plays in *globin* gene regulation could be further characterized. This work is the first time that Wnt has been shown to be involved in the regulation of *globin* gene expression, but does not identify the mechanism though which this regulation occurs. In order to determine the mechanism, ChIP-seq, RNA-seq, and knockdown experiments could be performed, similar to those used to elucidate a better mechanistic understanding of the thyroid hormone receptor pathway. As was initially determined in the knockdown screen, *tcf7l2* is involved in the regulation of *globin* gene expression, which can focus the initial investigation of the Wnt pathway. As a phenotype has been observed in both K562 and primary $CD34^+$ cells, ChIP-seq for TCF7L2 could be performed in these cell types. In K562 cells, the binding patterned could be observed and compared

between untreated, γ-globin expressing cells and T4 treated cells expressing higher levels of β-globin. The effect of XAV939 in cell culture could also be examined and, if a phenotype is observed, used as another comparison condition for ChIP-seq experiments. In $CD34^+$ cells, fetal and adult cells could be compared to determine if normal developmental differences exist for TCF7L2 binding as the globin genes change. Similarly, the various drug combinations could be compared to determine how binding patterns change under conditions that alter globin gene expression patterns. These conditions could also be compared using RNA-seq to determine which, if any, genes are altered. This would help to determine whether the Wnt pathway is acting directly on the globin locus or if it regulates the expression of other genes known to regulate globin gene expression. The result of these RNA-seq experiments could be combined with other information known about the functioning of the Wnt pathway, such as the roles of β-catenin, to perform knockdown/knockout experiments, either in zebrafish or cell culture, to determine which aspects of the Wnt pathway are responsible for its role in *globin* gene expression. These experiments could be aided by tools such as the zebrafish Wnt reporter line [257] and the CRISPR-Cas system for zebrafish [254] and human cells [252,253,255,256].

Once the role of the Wnt pathway is better understood, its interaction with the thyroid hormone receptor pathway could also be further examined. As critical factors are determined from the studies above and correlated with data gathered from this investigation of the thyroid hormone receptor pathway, co-immunoprecipitation experiments could be performed to determine if direct interactions between these pathways exist and are altered as changes in *globin* expression occur, either as the results of normal development or experimental manipulation. The results of the ChIP-seq and RNA-seq experiments above could be further examined to determine

if critical Wnt factors are binding to, and/or altering the transcription of, critical thyroid hormone receptor pathway members.

The role of the Wnt pathway in adult *globin* expressing states could also be further examined. The adult *globin* inducing role of T4 was identified in embryos, but further investigation demonstrated that inhibition of the pathway induced embryonic/larval *globin* expression in adults. Therefore, it is possible that, as inhibition of the Wnt pathway in embryos induces adult *globin*, activation of the Wnt pathway in adults would induce embryonic and/or larval *globin* expression. A Wnt activator such as BIO [258] that has been successfully used in zebrafish [145] could be injected retro-orbitally into adult zebrafish and the *globin* mRNA expression determined by RT-PCR and protein changes assessed with fluorescent antibody staining, analogous to the experiments completed for dronedarone here. The requirement of the Wnt pathway for the development and regeneration of the hematopoietic system, but its expendability for adult steady-state function [145], may indicate that the Wnt pathway generally promotes the embryonic-like state. It has been recently shown that the Wnt pathway, though TCF7L2 (the same factor identified here), reinforces lineage-specific regulators though direct binding of the factor to the same regions of the DNA, which then enhances expression and repression of the genes already targeted by the lineage regulator [145]. This is consistent with our results where knockdown of the Wnt pathway promotes an adult-like state. It suggests a model where loss of reinforcement of the embryonic state, loss of Wnt, allows lineage, and potentially stage, specific factors to have reduced influence and potentially allow "leakage" of incorrect gene expression. This would also be consistent with the low level induction observed with the Wnt inhibitors. This model could be used to further probe the role of the Wnt pathway

in *globin* gene regulation and potentially the distinction between the embryonic and adult states more generally.

As potential interactions between these two pathways could also exist in their role in promoting embryonic and/or fetal *globin* induction, they could be investigated using similar approaches as above. These interactions could be potentially advantageous for use in the clinic as a relatively large induction of fetal hemoglobin is needed to achieve a therapeutic threshold [176], and this could possibly be more easily achieved with a combination of agents from the Wnt and thyroid hormone receptor pathways.

Further mechanistic analysis of the thyroid hormone receptor pathway

While this work provides evidence to support the role of the thyroid hormone receptor pathway in the regulation of *globin* switching by direct action at the *globin* locus, further study is still needed to fully understand its role. The thyroid hormone receptors act in coordination with a number of heterodimeric partners, members of the RXR family, and coregulators, members of the Ncoa and Ncor families [150]. *Ncoa1* was initially identified in the knockdown screen performed here as suppressor of adult *globin* in the embryo. However, the exact role of the coregulator remains undetermined. In order to better assess the exact mechanism of the individual members of this pathway, it would be necessary to enrich for the affected cell population. As only a small fraction of the cells treated with either T4 or dronedarone alter their *globin* expression profile, assessment of the entire cell population will likely mask any changes occurring in the affected cells. However, due to technical limitations at this time, it is not possible to generate sufficient numbers of, or isolate and selectively interrogate, these cells. If this technical hurdle can be overcome, the role of the individual pathway members could be

elucidated though a combination of expression, protein interaction and ChIP-seq experiments. The expression status and level of each member of the pathway could be determined in induced embryonic, those expressing adult *globin*, and induced adult, those expressing embryonic *globin*, cells and compared to their untreated counterparts. This would set the landscape of which factors could be playing a role in the process, and provide potential hints as to which factors were important if significant expression changes were observed. The various complexes the factors form in the four conditions could be assessed using protein mass spectrometry, indicating which factors changes as a result of treatment. Finally, ChIP-seq experiments could be performed to verify the binding of these factors to known *globin* regulatory sites, implicating the alteration of certain proteins and protein complexes in the direct regulation of *globin* expression.

Interactions with known regulators

Neither the Wnt nor thyroid hormone receptor pathways are the only known regulators of *globin* switching. The interactions between these known pathways and regulators is essential to fully understanding the developmental process as well as more successfully pharmaceutically targeting one or more aspects for the treatment of hemoglobinopathies. The BCL11A pathway has emerged as a key repressor of fetal hemoglobin in humans and as a target for clinical modulation of globin expression [63,64,79]. The role of Bcl11a in *globin* switching has been investigated in zebrafish and exhibits similar functions in regulating *globin* switching as in other systems (Janelle Lambert, appendix). As the reagents and tools to interrogate this pathway exist in the zebrafish, including Bcl11a overexpression constructs, morpholinos for knockdown, and a genetic mutant from the Welcome Trust Sanger Institute [259] that exhibits a *globin* expression phenotype (David Wiley, personal communication), studies to determine the relationship between the Bcl11a pathway and the thyroid hormone receptor pathway could be studied in

zebrafish embryos and adults. By treating embryos with T4 and overexpressing Bcl11a, both of which increase the expression of adult *globin*, any additive, antagonistic, or synergistic effects could be assessed by RT-PCR and fluorescent antibody staining. This experiment would help to determine if and/or how these pathways work together. Conversely, embryos could be treated with combinations such as: (1) the Bcl11a morpholino and T4, (2) overexpressing Bcl11a and treating with dronedarone, and (3) the Bcl11a, Tcf7l2 and Ncoa1 morpholinos. Assessing the results of the opposing effects of these treatment pairs would help to begin to establish and define the epistasis interactions between these agents and pathways. Similarly, *bcl11a* mutant adult zebrafish could be treated with dronedarone, both of which should increase embryonic/larval *globin* expression, and the resulting combined effect assessed for additive, antagonistic, and synergistic effects. Additional combinations such as: (1) *bcl11a* mutants treated with T4 and (2) Bcl11a overexpressing fish treated with dronedarone, could be used to assess the epistatic relationships in the adult setting. Interesting results from these investigations could then be further expanded to include factors directly linked to the function of Bcl11a in its role as a regulator of hemoglobin switching, such as Sox6 [65]. These studies could also be expanded to include the interactions and epistatic interactions between the Wnt and Bcll1a pathways. Together, this work would aid in putting the data and observations made here into the broader context of the hemoglobin switching landscape, which could also further any potential goals of achieving clinical benefit for patients.

Assessment of genomic configuration

Looping of DNA from the locus control region (LCR) into physical interaction with the proximal *globin* gene promoters is a critical step in the regulation of *globin* expression by this essential enhancer of the region, and chromosome conformation capture techniques have been

invaluable in better understanding these physical interactions that regulate gene expression [35]. As the breath and variety of these techniques continues to expand, they can contribute even more to the general understanding of gene regulation, particularly at the *globin* locus [260]. As critical DNA elements have now been defined, with nearly base pair resolution, such as the major locus LCR and proximal *globin* promoters, these techniques are now feasible for use in the zebrafish to better understand both normal, developmental *globin* gene switching as well as thyroid hormone, Wnt, and Bcl11a pathway induced *globin* gene expression changes. These techniques can identify disparate regions of DNA that enter into close physical proximity to exhort regulatory control, as well as help to identify protein factors that cause, enhance, or otherwise promote these regulatory effects [260]. Incorporating an understanding of the physical interactions of the chromosome and the proteins that control them will significantly contribute to a better understanding of the *globin* switching process and role of pathways such as the thyroid hormone receptor pathway in these changes.

Interaction with mutants

One of the many advantages of the zebrafish system is the availability of a large number of mutants with hematological phenotypes [28,119]. These mutants can be used to help further characterize the *globin* gene expression phenotypes observed with modulation of the identified pathways. For example, the *zinfandel* mutant has a hypochromic microcytic anemia, but only during the embryonic stage [30]. As the phenotype of these mutants is correlated with the developmental stage of the erythrocytes, treatment of embryos with T4 or adults with dronedarone could yield results that would further support that the *globin* locus is specifically being altered by the thyroid hormone modulation, or that the entire state of the cell is being altered. As it is believed that the *zinfandel* mutation is located in a regulatory site upstream of

the major *globin* locus (data not shown, Yi Zhou, personal communication), further investigation of these pathways in this system could also help to establish which binding events these pathways alter are functional, and the particular function of individual binding sites.

Murine systems and clinical translation

While there are a large number of reasons why the zebrafish is an ideal system to study hemoglobin switching and cell culture models may be helpful in extrapolating the findings to other systems and opening up additional tool boxes, they have limitations when translating potential treatments to the clinic. In order to further investigate T4 modulation as a potential therapy, additional studies need to be completed in the murine system. However, mice have a single globin gene switch rather than the two observed in humans and zebrafish. Mice containing and expressing the entire human β-globin region in a developmentally regulated manner have been developed to study human globin regulation in a whole model organism [52]. This study has already shown that modulation of this pathway can alter globin gene expression in a mammalian system, but follow up studies in this model would be a critical step in translating this work into the clinic. As an effective dose (per gram of mass) and route of administration have already been established in the zebrafish, experimental procedures could be easily adapted to the murine system. With an expected time frame, type, and level of response known, efficient experiments could be designed to accommodate the more limited numbers of mice available compared to zebrafish. Preliminary data from further analysis of the data comparing various differentiation stages of fetal and adult human $CD34^+$ cells [261] shows that TRβ is one of the most differentially regulated genes, with significantly higher expression in the adult state (Jian Xu, personal communication). As differentiated fetal and adult $CD34^+$ cells expressed fetal and adult globin respectively, this difference may suggest that receptor expression, at least partially,

controls the effect of the thyroid hormone receptor pathway on globin expression and that TRβ is the thyroid hormone receptor that regulates globin switching in mammals. To investigate the role of TRβ further, the TRβ knockout mouse could be used. While no hematopoietic phenotype has been observed in the mouse [160–162], no work has been published examining *globin* switching. The lack of a hematopoietic phenotype may be advantageous as it may indicate that specific targeting of TRβ may limit side effects. Peripheral blood and bone marrow from knockout mice and wild type controls could be compared for expression of all globin genes at the mRNA and protein levels in the embryonic and adult stages to determine if the absence of this receptor alters normal globin expression. If a phenotype was observed, the mouse could be crossed into the transgenic mouse containing the human globin locus [52] to determine if the expression of the human globin genes are altered. If these results indicate that decreased levels of TRβ in the adult stage increase the expression of fetal globin, drugs specifically targeting TRβ could be tested in mouse models of sickle cell disease [102,103]. If these results then demonstrate that modulation of TRβ helps to affect the phenotype and/or ameliorate the symptoms of sickle cell disease, further translation of these drugs could be pursued for use in the clinical treatment of patients.

Conclusion

This thesis establishes the zebrafish as a model system for the study of hemoglobin switching and identified the Wnt and thyroid hormone receptor pathways as regulators of the process. By fully annotating all the *globin* genes on both *globin* loci and precisely determining the developmental timing of the two *globin* gene switches, it was possible to use the zebrafish to screen for agents that could alter the normal functioning of this process. After completing both a morpholino knockdown and chemical screen, two pathways were identified: the Wnt pathway

and the thyroid hormone receptor pathway. Further work showed that the thyroid hormone receptor pathway modulated both mRNA and protein expression in embryos and adults. Mammalian cell culture experiments demonstrated the conserved nature of this role, which potentially opens up opportunities to explore therapeutics for hemoglobinopathies. The techniques, reagents, and strategies developed here can be utilized in future screens and additional work to identify and investigate other aspects of *globin* gene regulation. This thesis can serve as the basis for the future study of hemoglobin switching in the zebrafish, the study of the Wnt and thyroid hormone receptor pathways in *globin* gene regulation, and as a potential new line of investigations in hemoglobinopathy therapies.

References

1 Korber, E. (1866) Uber Differengen des Blufarstoffs. Inaug. Diss. Univ. Dorpat.

2 Dzierzak, E. and Philipsen, S. (2013) Erythropoiesis: development and differentiation. Cold Spring Harb. Perspect. Med. **3**, a011601.

3 Hsia, N. and Zon, L. (2005) Transcriptional regulation of hematopoietic stem cell development in zebrafish. Exp. Hematol. **33**, 1007–1014.

4 McGrath, K. and Palis, J. (2008) Ontogeny of erythropoiesis in the mammalian embryo. Curr. Top. Dev. Biol. **82**, 1–22.

5 Bertrand, J. Y. and Traver, D. (2009) Hematopoietic cell development in the zebrafish embryo. Curr. Opin. Hematol. **16**, 243–248.

6 De Jong, J. L. O. and Zon, L. I. (2005) Use of the zebrafish system to study primitive and definitive hematopoiesis. Annu. Rev. Genet. **39**, 481–501.

7 Orkin, S. H. and Zon, L. I. (2008) Hematopoiesis: an evolving paradigm for stem cell biology. Cell **132**, 631–644.

8 Galloway, J. L. and Zon, L. I. (2003) Ontogeny of hematopoiesis: examining the emergence of hematopoietic cells in the vertebrate embryo. Curr. Top. Dev. Biol. **53**, 139–158.

9 Tsiftsoglou, A. S., Vizirianakis, I. S. and Strouboulis, J. (2009) Erythropoiesis: Model systems, molecular regulators, and developmental programs. IUBMB Life **61**, 800–830.

10 Palis, J. (2008) Ontogeny of erythropoiesis. Curr. Opin. Hematol. **15**, 155–161.

11 An, X. and Mohandas, N. (2011) Erythroblastic islands, terminal erythroid differentiation and reticulocyte maturation. Int. J. Hematol. **93**, 139–143.

12 Qian, F., Zhen, F., Xu, J., Huang, M., Li, W. and Wen, Z. (2007) Distinct functions for different scl isoforms in zebrafish primitive and definitive hematopoiesis. PLoS Biol. **5**, e132.

13 Katsumura, K. R., DeVilbiss, A. W., Pope, N. J., Johnson, K. D. and Bresnick, E. H. (2013) Transcriptional mechanisms underlying hemoglobin synthesis. Cold Spring Harb. Perspect. Med. **3**, a015412.

14 Kaneko, H., Shimizu, R. and Yamamoto, M. (2010) GATA factor switching during erythroid differentiation. Curr. Opin. Hematol. **17**, 163–168.

15 Bresnick, E. H., Katsumura, K. R., Lee, H.-Y., Johnson, K. D. and Perkins, A. S. (2012) Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. Nucleic Acids Res. **40**, 5819–5831.

16 Patterson, L. J., Gering, M., Eckfeldt, C. E., Green, A. R., Verfaillie, C. M., Ekker, S. C. and Patient, R. (2007) The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. Blood **109**, 2389–2398.

17 Kerenyi, M. A. and Orkin, S. H. (2010) Networking erythropoiesis. J. Exp. Med. **207**, 2537–2541.

18 Cantor, A. B. and Orkin, S. H. (2002) Transcriptional regulation of erythropoiesis: an affair involving multiple partners. Oncogene **21**, 3368–3376.

19 Bunn, H. F. (2013) Erythropoietin. Cold Spring Harb. Perspect. Med. **3**, a011619.

20 Hünefeld, F. L. (1840) Die Chemismus in der thierischen Organization 158–163.

21 Perutz, M. F., Fermi, G., Luisi, B., Shaanan, B. and Liddington, R. C. (1987) Stereochemistry of cooperative mechanisms in hemoglobin. Cold Spring Harb. Symp. Quant. Biol. **52**, 555–565.

22 Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. and North, A. C. (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature **185**, 416–422.

23 Ajioka, R. S., Phillips, J. D. and Kushner, J. P. (2006) Biosynthesis of heme in mammals. Biochim. Biophys. Acta **1763**, 723–736.

24 Broyles, R. H., Ramseyer, L. T., Do, T. H., McBride, K. A. and Barker, J. C. (1994) Hemoglobin switching in Rana/Xenopus erythroid heterokaryons: factors mediating the metamorphic hemoglobin switch are conserved. Dev. Genet. **15**, 347–355.

25 Hoffmann, F. G. and Storz, J. F. (2007) The D-Globin Gene Originated via Duplication of an Embryonic -Like Globin Gene in the Ancestor of Tetrapod Vertebrates. Mol. Biol. Evol. **24**, 1982–1990.

26 Kingsley, P. D. (2006) “Maturational” globin switching in primary primitive erythroid cells. Blood **107**, 1665–1672.

27 Chen, A. T. and Zon, L. I. (2009) Zebrafish blood stem cells. J. Cell. Biochem. **108**, 35–42.

28 Weinstein, B. M., Schier, A. F., Abdelilah, S., Malicki, J., Solnica-Krezel, L., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Driever, W. and Fishman, M. C. (1996) Hematopoietic mutations in the zebrafish. Dev. Camb. Engl. **123**, 303–309.

29 Haar, J. L. and Ackerman, G. A. (1971) A phase and electron microscopic study of vasculogenesis and erythropoiesis in the yolk sac of the mouse. Anat. Rec. **170**, 199–223.

30 Brownlie, A., Hersey, C., Oates, A. C., Paw, B. H., Falick, A. M., Witkowska, H. E., Flint, J., Higgs, D., Jessen, J. and Bahary, N. (2003) Characterization of embryonic globin genes of the zebrafish. Dev. Biol. **255**, 48–61.

31 Trimborn, T., Gribnau, J., Grosveld, F. and Fraser, P. (1999) Mechanisms of developmental control of transcription in the murine alpha - and beta -globin loci. Genes Dev. **13**, 112–124.

32 Chan, F. Y., Robinson, J., Brownlie, A., Shivdasani, R. A., Donovan, A., Brugnara, C., Kim, J., Lau, B. C., Witkowska, H. E. and Zon, L. I. (1997) Characterization of adult alpha- and beta-globin genes in the zebrafish. Blood **89**, 688–700.

33 Hardison, R. (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J. Exp. Biol. **201**, 1099–1117.

34 Gillemans, N., McMorrow, T., Tewari, R., Wai, A. W. K., Burgtorf, C., Drabek, D., Ventress, N., Langeveld, A., Higgs, D., Tan-Un, K., et al. (2003) Functional and comparative analysis of globin loci in pufferfish and humans. Blood **101**, 2842–2849.

35 Palstra, R.-J., de Laat, W. and Grosveld, F. (2008) Beta-globin regulation and long-range interactions. Adv. Genet. **61**, 107–142.

36 Li, Q., Peterson, K. R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus control regions. Blood **100**, 3077–3086.

37 Grosveld, F., van Assendelft, G. B., Greaves, D. R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell **51**, 975–985.

38 Noordermeer, D. and de Laat, W. (2008) Joining the loops: β-Globin gene regulation. IUBMB Life **60**, 824–833.

39 Goodwin, A. J., McInerney, J. M., Glander, M. A., Pomerantz, O. and Lowrey, C. H. (2001) In vivo formation of a human beta-globin locus control region core element requires binding sites for multiple factors including GATA-1, NF-E2, erythroid Kruppel-like factor, and Sp1. J. Biol. Chem. **276**, 26883–26892.

40 Higgs, D. R., Vernimmen, D. and Wood, B. (2008) Long-range regulation of alpha-globin gene expression. Adv. Genet. **61**, 143–173.

41 Anguita, E., Johnson, C. A., Wood, W. G., Turner, B. M. and Higgs, D. R. (2001) Identification of a conserved erythroid specific domain of histone acetylation across the alpha-globin gene cluster. Proc. Natl. Acad. Sci. U. S. A. **98**, 12114–12119.

42 Flint, J., Tufarelli, C., Peden, J., Clark, K., Daniels, R. J., Hardison, R., Miller, W., Philipsen, S., Tan-Un, K. C., McMorrow, T., et al. (2001) Comparative genome analysis

delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. Hum. Mol. Genet. **10**, 371–382.

43 Gumucio, D. L., Shelton, D. A., Zhu, W., Millinoff, D., Gray, T., Bock, J. H., Slightom, J. L. and Goodman, M. (1996) Evolutionary strategies for the elucidation of cis and trans factors that regulate the developmental switching programs of the beta-like globin genes. Mol. Phylogenet. Evol. **5**, 18–32.

44 Hardison, R. C. (2012) Evolution of hemoglobin and its genes. Cold Spring Harb. Perspect. Med. **2**, a011627.

45 Hughes, J. R. (2005) Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc. Natl. Acad. Sci. **102**, 9830–9835.

46 De Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. **11**, 447–459.

47 Bank, A. (2006) Regulation of human fetal hemoglobin: new players, new complexities. Blood **107**, 435–443.

48 Fang, X., Yin, W., Xiang, P., Han, H., Stamatoyannopoulos, G. and Li, Q. (2009) The higher structure of chromatin in the LCR of the beta-globin locus changes during development. J. Mol. Biol. **394**, 197–208.

49 Behringer, R. R., Ryan, T. M., Palmiter, R. D., Brinster, R. L. and Townes, T. M. (1990) Human gamma- to beta-globin gene switching in transgenic mice. Genes Dev. **4**, 380–389.

50 Enver, T., Ebens, A. J., Forrester, W. C. and Stamatoyannopoulos, G. (1989) The human beta-globin locus activation region alters the developmental fate of a human fetal globin gene in transgenic mice. Proc. Natl. Acad. Sci. U. S. A. **86**, 7033–7037.

51 Old, J. M. (2003) Screening and genetic diagnosis of haemoglobin disorders. Blood Rev. **17**, 43–53.

52 Porcu, S., Kitamura, M., Witkowska, E., Zhang, Z., Mutero, A., Lin, C., Chang, J. and Gaensler, K. M. (1997) The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. Blood **90**, 4602–4609.

53 Kukreti, S., Kaur, H., Kaushik, M., Bansal, A., Saxena, S., Kaushik, S. and Kukreti, R. (2010) Structural polymorphism at LCR and its role in beta-globin gene regulation. Biochimie.

54 Kulozik, A. E., Yarwood, N. and Jones, R. W. (1988) The Corfu delta beta zero thalassemia: a small deletion acts at a distance to selectively abolish beta globin gene expression. Blood **71**, 457–462.

55 Broyles, R. H., Johnson, G. M., Maples, P. B. and Kindell, G. R. (1981) Two erythropietic microenvironments and two larval red cell lines in bullfrog tadpoles. Dev. Biol. **81**, 299–314.

56 Ramseyer, L. T., Barker-Harrel, J., Smith, D. J., McBride, K. A., Jarman, R. N. and Broyles, R. H. (1989) Intracellular signals for developmental hemoglobin switching. Dev. Biol. **133**, 262–271.

57 Broyles, R. H. (1999) Use of somatic cell fusion to reprogram globin genes. Semin. Cell Dev. Biol. **10**, 259–265.

58 Broyles, R. H., Roth, A. C., Todd, M. and Belegu, V. (2006) Nuclear reprogramming by cell fusion. Methods Mol. Biol. Clifton NJ **325**, 47–57.

59 Brinkman, R., Wildschut, A. and Wittermans, A. (1934) On the occurrence of two kinds of haemoglobin in normal human blood. J. Physiol. **80**, 377–387.

60 Lettre, G., Sankaran, V. G., Bezerra, M. A. C., Araújo, A. S., Uda, M., Sanna, S., Cao, A., Schlessinger, D., Costa, F. F., Hirschhorn, J. N., et al. (2008) DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl. Acad. Sci. U. S. A. **105**, 11869–11874.

61 Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007) A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat. Genet. **39**, 1197–1199.

62 Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V. G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008) Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci. U. S. A. **105**, 1620–1625.

63 Sankaran, V. G., Menne, T. F., Xu, J., Akie, T. E., Lettre, G., Van Handel, B., Mikkola, H. K. A., Hirschhorn, J. N., Cantor, A. B. and Orkin, S. H. (2008) Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science **322**, 1839–1842.

64 Sankaran, V. G., Xu, J., Ragoczy, T., Ippolito, G. C., Walkley, C. R., Maika, S. D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M. A., et al. (2009) Developmental and species-divergent globin switching are driven by BCL11A. Nature **460**, 1093–1097.

65 Xu, J., Sankaran, V. G., Ni, M., Menne, T. F., Puram, R. V., Kim, W. and Orkin, S. H. (2010) Transcriptional silencing of -globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. **24**, 783–798.

66 Xu, J., Peng, C., Sankaran, V. G., Shao, Z., Esrick, E. B., Chong, B. G., Ippolito, G. C., Fujiwara, Y., Ebert, B. L., Tucker, P. W., et al. (2011) Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science **334**, 993–996.

67 Bauer, D. E., Kamran, S. C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M. C., Smith, E. C., Pinello, L., et al. (2013) An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science **342**, 253–257.

68 Sankaran, V. G. and Orkin, S. H. (2013) The switch from fetal to adult hemoglobin. Cold Spring Harb. Perspect. Med. **3**, a011643.

69 Xu, J., Bauer, D. E., Kerenyi, M. A., Vo, T. D., Hou, S., Hsu, Y.-J., Yao, H., Trowbridge, J. J., Mandel, G. and Orkin, S. H. (2013) Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. U. S. A. **110**, 6518–6523.

70 Donze, D., Townes, T. M. and Bieker, J. J. (1995) Role of erythroid Kruppel-like factor in human gamma- to beta-globin gene switching. J. Biol. Chem. **270**, 1955–1959.

71 Tallack, M. R. and Perkins, A. C. (2013) Three fingers on the switch: Krüppel-like factor 1 regulation of γ-globin to β-globin gene switching. Curr. Opin. Hematol. **20**, 193–200.

72 Song, S.-H., Hou, C. and Dean, A. (2007) A Positive Role for NLI/Ldb1 in Long-Range β-Globin Locus Control Region Function. Mol. Cell **28**, 810–822.

73 Song, S.-H., Kim, A., Ragoczy, T., Bender, M. A., Groudine, M. and Dean, A. (2010) Multiple functions of Ldb1 required for β-globin activation during erythroid differentiation. Blood.

74 Bell, A. C., West, A. G. and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell **98**, 387–396.

75 Bulger, M., Schübeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C. and Groudine, M. (2003) A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. Mol. Cell. Biol. **23**, 5234–5244.

76 Sankaran, V. G., Menne, T. F., Scepanovic, D., Vergilio, J.-A., Ji, P., Kim, J., Thiru, P., Orkin, S. H., Lander, E. S. and Lodish, H. F. (2011) MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. Proc. Natl. Acad. Sci. U. S. A.

77 Keys, J. R., Tallack, M. R., Zhan, Y., Papathanasiou, P., Goodnow, C. C., Gaensler, K. M., Crossley, M., Dekker, J. and Perkins, A. C. (2008) A mechanism for Ikaros regulation of human globin gene switching. Br. J. Haematol. 080305033838221–???

78 Lopez, R. A. (2002) Multiple hematopoietic defects and delayed globin switching in Ikaros null mice. Proc. Natl. Acad. Sci. **99**, 602–607.

79 Sankaran, V. G. (2011) Targeted therapeutic strategies for fetal hemoglobin induction. Hematol. Educ. Program Am. Soc. Hematol. Am. Soc. Hematol. Educ. Program **2011**, 459–465.

80 Honig, R., Billing, R., Matsuyama, M. and Terasaki, P. J. (1976) B lymphocyte alloantigen specificities present on cultured lymphoblastoid cell lines. J. Immunol. Baltim. Md 1950 **117**, 1300–1302.

81 Kioussis, D., Wilson, F., Khazaie, K. and Grosveld, F. (1985) Differential expression of human globin genes introduced in K562 cells. EMBO J. **4**, 927–931.

82 Lozzio, C. B. and Lozzio, B. B. (1975) Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. Blood **45**, 321–334.

83 Brown, B. A., Padgett, R. W., Hardies, S. C., Hutchison, C. A., 3rd and Edgell, M. H. (1982) beta-globin transcript found in induced murine erythroleukemia cells is homologous to the beta h0 and beta h1 genes. Proc. Natl. Acad. Sci. U. S. A. **79**, 2753–2757.

84 Ikawa, Y., Aida, M. and Inoue, Y. (1976) Isolation and characterization of high and low differentiation-inducible Friend leukemia lines. Gann Gan **67**, 767–770.

85 Singer, D., Cooper, M., Maniatis, G. M., Marks, P. A. and Rifkind, R. A. (1974) Erythropoietic differentiation in colonies of cells transformed by Friend virus. Proc. Natl. Acad. Sci. U. S. A. **71**, 2668–2670.

86 Andrews, R. G., Singer, J. W. and Bernstein, I. D. (1986) Monoclonal antibody 12-8 recognizes a 115-kd molecule present on both unipotent and multipotent hematopoietic colony-forming cells and their precursors. Blood **67**, 842–845.

87 Civin, C. I., Strauss, L. C., Brovall, C., Fackler, M. J., Schwartz, J. F. and Shaper, J. H. (1984) Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. J. Immunol. Baltim. Md 1950 **133**, 157–165.

88 Tindle, R. W., Nichols, R. A., Chan, L., Campana, D., Catovsky, D. and Birnie, G. D. (1985) A novel monoclonal antibody BI-3C5 recognises myeloblasts and non-B non-T lymphoblasts in acute leukaemias and CGL blast crises, and reacts with immature cells in normal bone marrow. Leuk. Res. **9**, 1–9.

89 Watt, S. M., Karhi, K., Gatter, K., Furley, A. J., Katz, F. E., Healy, L. E., Altass, L. J., Bradley, N. J., Sutherland, D. R. and Levinsky, R. (1987) Distribution and epitope analysis of the cell membrane glycoprotein (HPCA-1) associated with human hemopoietic progenitor cells. Leukemia **1**, 417–426.

90 Olivier, E., Qiu, C. and Bouhassira, E. E. (2012) Novel, high-yield red blood cell production methods from CD34-positive cells derived from human embryonic stem, yolk sac, fetal liver, cord blood, and peripheral blood. Stem Cells Transl. Med. **1**, 604–614.

91 Baglioni, C. and Sparks, C. E. (1963) A Study of Hemoglobin Differentiation in Rana Catesbeiana. Dev. Biol. **8**, 272–285.

92 Trader, C. D., Wortham, J. S. and Frieden, E. (1963) Hemoglobin: molecular changes during anuran metamorphosis. Science **139**, 918–919.

93 Flavin, M., Blouquit, Y. and Rosa, J. (1978) Biochemical studies of the hemoglobin switch during metamorphosis in the salamander Pleurodeles waltli. I. Partial characterization of the adult hemoglobin. Comp. Biochem. Physiol. **61 B**, 533–537.

94 Flavin, M., Blouquit, Y., Duprat, A. M. and Rosa, J. (1978) Biochemical studies of the hemoglobin switch during metamorphosis in the salamander Pleurodeles waltli. II. Comparative studies of larval and adult hemoglobins. Comp. Biochem. Physiol. **61 B**, 539–544.

95 Fantoni, A., Bank, A. and Marks, P. A. (1967) Globin composition and synthesis of hemoglobins in developing fetal mice erythroid cells. Science **157**, 1327–1329.

96 Leder, P., Hansen, J. N., Konkel, D., Leder, A., Nishioka, Y. and Talkington, C. (1980) Mouse globin system: a functional and evolutionary analysis. Science **209**, 1336–1342.

97 Russell, E. S. and McFarland, E. C. (1974) Genetics of mouse hemoglobins. Ann. N. Y. Acad. Sci. **241**, 25–38.

98 Stamatoyannopoulos, G. (2005) Control of globin gene expression during development and erythroid differentiation. Exp. Hematol. **33**, 259–271.

99 Higgs, D. R., Wood, W. G., Jarman, A. P., Sharpe, J., Lida, J., Pretorius, I. M. and Ayyub, H. (1990) A major positive regulatory region located far upstream of the human alpha-globin gene locus. Genes Dev. **4**, 1588–1601.

100 Palmiter, R. D. and Brinster, R. L. (1986) Germ-line transformation of mice. Annu. Rev. Genet. **20**, 465–499.

101 Peterson, K. R., Navas, P. A. and Stamatoyannopoulos, G. (1998) beta-YAC transgenic mice for studying LCR function. Ann. N. Y. Acad. Sci. **850**, 28–37.

102 Pászty, C., Brion, C. M., Manci, E., Witkowska, H. E., Stevens, M. E., Mohandas, N. and Rubin, E. M. (1997) Transgenic knockout mice with exclusively human sickle hemoglobin and sickle cell disease. Science **278**, 876–878.

103 Ryan, T. M., Ciavatta, D. J. and Townes, T. M. (1997) Knockout-transgenic mouse model of sickle cell disease. Science **278**, 873–876.

104 Herrick, J. B. (1910) Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. Arch. Intern. Med. **6**, 517–65.

105 Cooley, T., Witwer, E. and Lee, P. (1927) Anemia in children: With splenomegaly and peculiar changes in the bones report of cases. Am. J. Dis. Child. **34**, 347–363.

106 Whipple, G. and Bradford, W. (1932) Racial or familial anemia of children: Associated with fundamental disturbances of bone and pigment metabolism (cooley-von jaksch). Am. J. Dis. Child. **44**, 336–365.

107 Chakalova, L., Osborne, C. S., Dai, Y.-F., Goyenechea, B., Metaxotou-Mavromati, A., Kattamis, A., Kattamis, C. and Fraser, P. (2005) The Corfu deltabeta thalassemia deletion disrupts gamma-globin gene silencing and reveals post-transcriptional regulation of HbF expression. Blood **105**, 2154–2160.

108 Jing, L. and Zon, L. I. (2011) Zebrafish as a model for normal and malignant hematopoiesis. Dis. Model. Mech. **4**, 433–438.

109 Bowman, T. V. and Zon, L. I. (2010) Swimming into the future of drug discovery: in vivo chemical screens in zebrafish. ACS Chem. Biol. **5**, 159–161.

110 Huang, H.-T., Kathrein, K. L., Barton, A., Gitlin, Z., Huang, Y.-H., Ward, T. P., Hofmann, O., Dibiase, A., Song, A., Tyekucheva, S., et al. (2013) A network of epigenetic regulators guides developmental haematopoiesis in vivo. Nat. Cell Biol. **15**, 1516–1525.

111 Jing, L., Durand, E. M., Ezzio, C., Pagliuca, S. M. and Zon, L. I. (2012) In situ hybridization assay-based small molecule screening in zebrafish. Curr. Protoc. Chem. Biol. **4**.

112 Kaufman, C. K., White, R. M. and Zon, L. (2009) Chemical genetic screening in the zebrafish embryo. Nat. Protoc. **4**, 1422–1432.

113 Ceol, C. J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D. A., Battisti, V., Fritsch, L., Lin, W. M., Hollmann, T. J., Ferré, F., et al. (2011) The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. Nature **471**, 513–517.

114 Taylor, J. S., Braasch, I., Frickey, T., Meyer, A. and Van de Peer, Y. (2003) Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. **13**, 382–390.

115 Henn, K. and Braunbeck, T. (2011) Dechorionation as a tool to improve the fish embryo toxicity test (FET) with the zebrafish (Danio rerio). Comp. Biochem. Physiol. Toxicol. Pharmacol. CBP **153**, 91–98.

116 Orkin, S. H. and Zon, L. I. (1997) Genetics of erythropoiesis: induced mutations in mice and zebrafish. Annu. Rev. Genet. **31**, 33–60.

117 Burns, C. E., Galloway, J. L., Smith, A. C. H., Keefe, M. D., Cashman, T. J., Paik, E. J., Mayhall, E. A., Amsterdam, A. H. and Zon, L. I. (2009) A genetic screen in zebrafish defines a hierarchical network of pathways required for hematopoietic stem cell emergence. Blood **113**, 5776–5782.

118 Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. Dev. Camb. Engl. **123**, 1–36.

119 Ransom, D. G., Haffter, P., Odenthal, J., Brownlie, A., Vogelsang, E., Kelsh, R. N., Brand, M., van Eeden, F. J., Furutani-Seiki, M., Granato, M., et al. (1996) Characterization of zebrafish mutants with defects in embryonic hematopoiesis. Dev. Camb. Engl. **123**, 311–319.

120 Grillitsch, S., Medgyesy, N., Schwerte, T. and Pelster, B. (2005) The influence of environmental P(O(2)) on hemoglobin oxygen saturation in developing zebrafish Danio rerio. J. Exp. Biol. **208**, 309–316.

121 Pelster, B. and Burggren, W. W. (1996) Disruption of hemoglobin oxygen transport does not impact oxygen-dependent physiological processes in developing embryos of zebra fish (Danio rerio). Circ. Res. **79**, 358–362.

122 Long, Q., Meng, A., Wang, H., Jessen, J. R., Farrell, M. J. and Lin, S. (1997) GATA-1 expression pattern can be recapitulated in living transgenic zebrafish using GFP reporter gene. Dev. Camb. Engl. **124**, 4105–4111.

123 Kwan, K. M., Fujimoto, E., Grabher, C., Mangum, B. D., Hardy, M. E., Campbell, D. S., Parant, J. M., Yost, H. J., Kanki, J. P. and Chien, C.-B. (2007) The Tol2kit: a multisite gateway-based construction kit forTol2 transposon transgenesis constructs. Dev. Dyn. **236**, 3088–3099.

124 Postlethwait, J. H., Yan, Y. L., Gates, M. A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E. S., Force, A., Gong, Z., et al. (1998) Vertebrate genome evolution and the zebrafish gene map. Nat. Genet. **18**, 345–349.

125 Quinkertz, A. and Campos-Ortega, J. A. (1999) A new beta-globin gene from the zebrafish, betaE1, and its pattern of transcription during embryogenesis. Dev. Genes Evol. **209**, 126–131.

126 Maruyama, K., Yasumasu, S., Naruse, K., Mitani, H., Shima, A. and Iuchi, I. (2004) Genomic organization and developmental expression of globin genes in the teleost Oryzias latipes. Gene **335**, 89–100.

127 Tiedke, J., Gerlach, F., Mitz, S. A., Hankeln, T. and Burmester, T. (2011) Ontogeny of globin expression in zebrafish (Danio rerio). J. Comp. Physiol. [B] **181**, 1011–1021.

128 Roesner, A. (2006) Hypoxia induces a complex response of globin expression in zebrafish (Danio rerio). J. Exp. Biol. **209**, 2129–2137.

129 Wawrowski, A., Gerlach, F., Hankeln, T. and Burmester, T. (2011) Changes of globin expression in the Japanese medaka (Oryzias latipes) in response to acute and chronic hypoxia. J. Comp. Physiol. [B] **181**, 199–208.

130 Risso, A., Fabbro, D., Damante, G. and Antonutto, G. (2012) Expression of fetal hemoglobin in adult humans exposed to high altitude hypoxia. Blood Cells. Mol. Dis. **48**, 147–153.

131 Salomon-Andonie, J., Miasnikova, G., Sergueeva, A., Polyakova, L. A., Niu, X., Nekhai, S. and Gordeuk, V. R. (2013) Effect of congenital upregulation of hypoxia inducible factors on percentage of fetal hemoglobin in the blood. Blood **122**, 3088–3089.

132 Gardiner, M. R., Gongora, M. M., Grimmond, S. M. and Perkins, A. C. (2007) A global role for zebrafish klf4 in embryonic erythropoiesis. Mech. Dev. **124**, 762–774.

133 Kawahara, A. and Dawid, I. B. (2001) Critical role of biklf in erythroid cell differentiation in zebrafish. Curr. Biol. CB **11**, 1353–1357.

134 Oates, A. C. (2001) The zebrafish klf gene family. Blood **98**, 1792–1801.

135 Fu, Y.-F., Du, T.-T., Dong, M., Zhu, K.-Y., Jing, C.-B., Zhang, Y., Wang, L., Fan, H.-B., Chen, Y., Jin, Y., et al. (2009) Mir-144 selectively regulates embryonic -hemoglobin synthesis during primitive erythropoiesis. Blood **113**, 1340–1349.

136 Buechling, T. and Boutros, M. (2011) Chapter two - Wnt Signaling: Signaling at and Above the Receptor Level. In Current Topics in Developmental Biology (Carmen Birchmeier, ed.), pp 21–53, Academic Press.

137 Clevers, H. and Nusse, R. (2012) Wnt/β-Catenin Signaling and Disease. Cell **149**, 1192–1205.

138 MacDonald, B. T., Tamai, K. and He, X. (2009) Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. Dev. Cell **17**, 9–26.

139 Niehrs, C. (2012) The complex world of WNT receptor signalling. Nat. Rev. Mol. Cell Biol. **13**, 767–779.

140 Staal, F. J. T. and Clevers, H. C. (2005) WNT signalling and haematopoiesis: a WNT-WNT situation. Nat. Rev. Immunol. **5**, 21–30.

141 Wang, J., Sinha, T. and Wynshaw-Boris, A. (2012) Wnt signaling in mammalian development: lessons from mouse genetics. Cold Spring Harb. Perspect. Biol. **4**.

142 Nemeth, M. J. and Bodine, D. M. (2007) Regulation of hematopoiesis and the hematopoietic stem cell niche by Wnt signaling pathways. Cell Res. **17**, 746–758.

143 Staal, F. J. T. and Luis, T. C. (2010) Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. J. Cell. Biochem. **109**, 844–849.

144 Wilusz, M. and Majka, M. (2008) Role of the Wnt/beta-catenin network in regulating hematopoiesis. Arch. Immunol. Ther. Exp. (Warsz.) **56**, 257–266.

145 Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D.-C., DiBiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L., et al. (2011) Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell **147**, 577–589.

146 Gronemeyer, H., Gustafsson, J.-A. and Laudet, V. (2004) Principles for modulation of the nuclear receptor superfamily. Nat. Rev. Drug Discov. **3**, 950–964.

147 Oetting, A. and Yen, P. M. (2007) New insights into thyroid hormone action. Best Pract. Res. Clin. Endocrinol. Metab. **21**, 193–208.

148 Tarım, Ö. (2011) Thyroid hormones and growth in health and disease. J. Clin. Res. Pediatr. Endocrinol. **3**, 51–55.

149 Bianco, A. C. and Kim, B. W. (2006) Deiodinases: implications of the local control of thyroid hormone action. J. Clin. Invest. **116**, 2571–2579.

150 Astapova, I. and Hollenberg, A. N. (2012) The in vivo role of nuclear receptor corepressors in thyroid hormone action. Biochim. Biophys. Acta.

151 Wu, Y., Xu, B. and Koenig, R. J. (2001) Thyroid hormone response element sequence and the recruitment of retinoid X receptors for thyroid hormone responsiveness. J. Biol. Chem. **276**, 3929–3936.

152 Bertrand, S., Thisse, B., Tavares, R., Sachs, L., Chaumot, A., Bardet, P.-L., Escrivà, H., Duffraisse, M., Marchand, O., Safi, R., et al. (2007) Unexpected novel relational links uncovered by extensive developmental profiling of nuclear receptor expression. PLoS Genet. **3**, e188.

153 Grimaldi, A., Buisine, N., Miller, T., Shi, Y.-B. and Sachs, L. M. (2013) Mechanisms of thyroid hormone receptor action during development: lessons from amphibian studies. Biochim. Biophys. Acta **1830**, 3882–3892.

154 Swingle, W. W. (1919) Iodine and the thyroid : III. The specific action of iodine in accelerating amphibian metamorphosis. J. Gen. Physiol. **1**, 593–606.

155 McMenamin, S. K. and Parichy, D. M. (2013) Metamorphosis in teleosts. Curr. Top. Dev. Biol. **103**, 127–165.

156 Chang, J., Wang, M., Gui, W., Zhao, Y., Yu, L. and Zhu, G. (2012) Changes in thyroid hormone levels during zebrafish development. Zoolog. Sci. **29**, 181–184.

157 Forman, L. J. and Just, J. J. (1981) Cellular quantitation of hemoglobin transition during natural and thyroid-hormone-induced metamorphosis of the bullfrog, Rana catesbeiana. Gen. Comp. Endocrinol. **44**, 1–12.

158 Mukhi, S., Cai, L. and Brown, D. D. (2010) Gene switching at Xenopus laevis metamorphosis. Dev. Biol. **338**, 117–126.

159 Holzer, G. and Laudet, V. (2013) Thyroid hormones and postembryonic development in amniotes. Curr. Top. Dev. Biol. **103**, 397–425.

160 Flamant, F. and Samarut, J. (2003) Thyroid hormone receptors: lessons from knockout and knock-in mutant mice. Trends Endocrinol. Metab. TEM **14**, 85–90.

161 Hsu, J. H. and Brent, G. A. (1998) Thyroid hormone receptor gene knockouts. Trends Endocrinol. Metab. TEM **9**, 103–112.

162 O'Shea, P. J. and Williams, G. R. (2002) Insight into the physiological actions of thyroid hormone receptors from genetically modified mice. J. Endocrinol. **175**, 553–570.

163 Angelin-Duclos, C., Domenget, C., Kolbus, A., Beug, H., Jurdic, P. and Samarut, J. (2005) Thyroid hormone T3 acting through the thyroid hormone alpha receptor is necessary for implementation of erythropoiesis in the neonatal spleen environment in the mouse. Dev. Camb. Engl. **132**, 925–934.

164 Kendrick, T. S., Payne, C. J., Epis, M. R., Schneider, J. R., Leedman, P. J., Klinken, S. P. and Ingley, E. (2008) Erythroid defects in TRalpha-/- mice. Blood **111**, 3245–3248.

165 Beildeck, M. E., Gelmann, E. P. and Byers, S. W. (2010) Cross-regulation of signaling pathways: an example of nuclear hormone receptors and the canonical Wnt pathway. Exp. Cell Res. **316**, 1763–1772.

166 Mulholland, D. J., Dedhar, S., Coetzee, G. A. and Nelson, C. C. (2005) Interaction of nuclear receptors with the Wnt/beta-catenin/Tcf signaling axis: Wnt you like to know? Endocr. Rev. **26**, 898–915.

167 Plateroti, M., Kress, E., Mori, J. I. and Samarut, J. (2006) Thyroid hormone receptor alpha1 directly controls transcription of the beta-catenin gene in intestinal epithelial cells. Mol. Cell. Biol. **26**, 3204–3214.

168 Sirakov, M., Skah, S., Lone, I. N., Nadjar, J., Angelov, D. and Plateroti, M. (2012) Multi-level interactions between the nuclear receptor TRα1 and the WNT effectors β-catenin/Tcf4 in the intestinal epithelium. PloS One **7**, e34162.

169 Ganis, J. J., Hsia, N., Trompouki, E., de Jong, J. L. O., DiBiase, A., Lambert, J. S., Jia, Z., Sabo, P. J., Weaver, M., Sandstrom, R., et al. (2012) Zebrafish globin switching occurs in two developmental stages and is controlled by the LCR. Dev. Biol. **366**, 185–194.

170 Li, Q., Peterson, K. R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus control regions. Blood **100**, 3077–3086.

171 Sankaran, V. G., Xu, J. and Orkin, S. H. (2010) Advances in the understanding of haemoglobin switching. Br. J. Haematol. **149**, 181–194.

172 Johnson, R. M., Gumucio, D. and Goodman, M. (2002) Globin gene switching in primates. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. **133**, 877–883.

173 Galanello, R. and Origa, R. (2010) Beta-thalassemia. Orphanet J. Rare Dis. **5**, 11.

174 Harteveld, C. L. and Higgs, D. R. (2010) Alpha-thalassaemia. Orphanet J. Rare Dis. **5**, 13.

175 Mousa, S. A. and Qari, M. H. (2010) Diagnosis and management of sickle cell disorders. Methods Mol. Biol. Clifton NJ **663**, 291–307.

176 Orkin, S. H. and Higgs, D. R. (2010) Sickle cell disease at 100 years. Science **329**, 291–292.

177 World Health Organization and Thalassemia International Foundation. (2008) Management of Haemoglobin Disorders: Report of Joint WHO-TIF Meeting, Nicosia, Cyprus, World Health Organization, Geneva, Switzerland.

178 Eaton, W. A. (2003) Linus Pauling and sickle cell disease. Biophys. Chem. **100**, 109–116.

179 Pauling, L., Itano, H., Singer, S. J. and Wells, I. C. (1949) Sickle Cell Anemia, a Molecular Disease. Science **110**, 543–548.

180 Fritsch, E. F., Lawn, R. M. and Maniatis, T. (1980) Molecular cloning and characterization of the human beta-like globin gene cluster. Cell **19**, 959–972.

181 Watson, J., Starman, A. W. and Bilello, F. P. (1948) The significance of the paucity of sickle cells in newborn negro infants. Am. J. Med. Sci **215**, 419–423.

182 Higgs, D. R. and Wood, W. G. (2008) Long-range regulation of alpha globin gene expression during erythropoiesis. Curr. Opin. Hematol. **15**, 176–183.

183 Goodman, M., Moore, G. W. and Matsuda, G. (1975) Darwinian evolution in the genealogy of haemoglobin. Nature **253**, 603–608.

184 Bahary, N. and Zon, L. I. (1998) Use of the zebrafish (Danio rerio) to define hematopoiesis. Stem Cells Dayt. Ohio **16**, 89–98.

185 Shafizadeh, E. and Paw, B. H. (2004) Zebrafish as a model of human hematologic disorders. Curr. Opin. Hematol. **11**, 255–261.

186 Rombough, P. and Drader, H. (2009) Hemoglobin enhances oxygen uptake in larval zebrafish (Danio rerio) but only under conditions of extreme hypoxia. J. Exp. Biol. **212**, 778–784.

187 Trompouki, E. and Zon, L. I. (2010) Small molecule screen in zebrafish and HSC expansion. Methods Mol. Biol. Clifton NJ **636**, 301–316.

188 Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. and Schilling, T. F. (1995) Stages of embryonic development of the zebrafish. Dev. Dyn. Off. Publ. Am. Assoc. Anat. **203**, 253–310.

189 Westerfield, M. (2000) The Zebrafish Book: A guide for the laboratory use of zebrafish (Danio rerio). 4th ed. 4th ed., University of Oregon Press, Eugene, OR.

190 Lin, H.-F. (2005) Analysis of thrombocyte development in CD41-GFP transgenic zebrafish. Blood **106**, 3803–3810.

191 Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**, R25.

192 Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol. **9**, R137.

193 Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat. Methods **3**, 511–518.

194 Hesselberth, J. R., Chen, X., Zhang, Z., Sabo, P. J., Sandstrom, R., Reynolds, A. P., Thurman, R. E., Neph, S., Kuehn, M. S., Noble, W. S., et al. (2009) Global mapping of

protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods **6**, 283–289.

195 Lee, T. I., Johnstone, S. E. and Young, R. A. (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. **1**, 729–748.

196 Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E. and Young, R. A. (2008) Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev. **22**, 3403–3408.

197 Thisse, C. and Thisse, B. (2008) High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc. **3**, 59–69.

198 Edgar, R., Domrachev, M. and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30**, 207–210.

199 Higgs, D. R., Garrick, D., Anguita, E., De Gobbi, M., Hughes, J., Muers, M., Vernimmen, D., Lower, K., Law, M., Argentaro, A., et al. (2005) Understanding alpha-globin gene regulation: Aiming to improve the management of thalassemia. Ann. N. Y. Acad. Sci. **1054**, 92–102.

200 Giles, K. E., Gowher, H., Ghirlando, R., Jin, C. and Felsenfeld, G. (2010) Chromatin Boundaries, Insulators, and Long-Range Interactions in the Nucleus. Cold Spring Harb. Symp. Quant. Biol. **75**, 79–85.

201 Maruyama, K., Ishikawa, Y., Yasumasu, S. and Iuchi, I. (2007) Globin gene enhancer activity of a DNase-I hypersensitive site-40 homolog in medaka, Oryzias latipes. Zoolog. Sci. **24**, 997–1004.

202 Bertrand, J. Y., Kim, A. D., Violette, E. P., Stachura, D. L., Cisson, J. L. and Traver, D. (2007) Definitive hematopoiesis initiates through a committed erythromyeloid progenitor in the zebrafish embryo. Development **134**, 4147–4156.

203 Murayama, E., Kissa, K., Zapata, A., Mordelet, E., Briolat, V., Lin, H.-F., Handin, R. I. and Herbomel, P. (2006) Tracing Hematopoietic Precursor Migration to Successive Hematopoietic Organs during Zebrafish Development. Immunity **25**, 963–975.

204 Enver, T., Raich, N., Ebens, A. J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990) Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. Nature **344**, 309–313.

205 Tang, Y., Wang, Z., Huang, Y., Liu, D., Liu, G., Shen, W., Tang, X., Feng, D. and Liang, C. (2006) Gene order in human α-globin locus is required for their temporal specific expressions. Genes Cells **11**, 123–131.

206 Guerrero, G., Delgadoolguin, P., Escamilladelarenal, M., Furlanmagaril, M., Rebollar, E., Delarosavelazquez, I., Sotoreyes, E., Rinconarano, H., Valdesquezada, C. and Valadezgraham, V. (2007) Globin genes transcriptional switching, chromatin structure and linked lessons to epigenetics in cancer: A comparative overview☆. Comp. Biochem. Physiol. - Part Mol. Integr. Physiol. **147**, 750–760.

207 Vernimmen, D., Marques-Kranc, F., Sharpe, J. A., Sloane-Stanley, J. A., Wood, W. G., Wallace, H. A. C., Smith, A. J. H. and Higgs, D. R. (2009) Chromosome looping at the human -globin locus is mediated via the major upstream regulatory element (HS -40). Blood **114**, 4253–4260.

208 Maniatis, G. M. and Ingram, V. M. (1971) Erythropoiesis during amphibian metamorphosis : I. Site of Maturation of Erythrocytes in Rana catesbeiana. J. Cell Biol. **49**, 372–379.

209 Maruyama, K., Yasumasu, S. and Iuchi, I. (2002) Characterization and expression of embryonic and adult globins of the teleost Oryzias latipes (medaka). J. Biochem. (Tokyo) **132**, 581–589.

210 Dailey, H. A. and Meissner, P. N. (2013) Erythroid heme biosynthesis and its disorders. Cold Spring Harb. Perspect. Med. **3**, a011676.

211 Perumbeti, A. and Malik, P. (2010) Therapy for beta-globinopathies: a brief review and determinants for successful and safe correction. Ann. N. Y. Acad. Sci. **1202**, 36–44.

212 Weatherall, D. J. (2013) The role of the inherited disorders of hemoglobin, the first "molecular diseases," in the future of human genetics. Annu. Rev. Genomics Hum. Genet. **14**, 1–24.

213 Zitnik, G., Hines, P., Stamatoyannopoulos, G. and Papayannopoulou, T. (1991) Murine erythroleukemia cell line GM979 contains factors that can activate silent chromosomal human gamma-globin genes. Proc. Natl. Acad. Sci. U. S. A. **88**, 2530–2534.

214 MacDonald, B. T., Tamai, K. and He, X. (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. Dev. Cell **17**, 9–26.

215 Morvan-Dubois, G., Fini, J. B. and Demeneix, B. A. (2013) Is thyroid hormone signaling relevant for vertebrate embryogenesis? Curr. Top. Dev. Biol. **103**, 365–396.

216 Kingsley, P. D., Greenfest-Allen, E., Frame, J. M., Bushnell, T. P., Malik, J., McGrath, K. E., Stoeckert, C. J. and Palis, J. (2013) Ontogeny of erythroid gene expression. Blood **121**, e5–e13.

217 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25**, 25–29.

218 Gan, X., Wang, J., Xi, Y., Wu, Z., Li, Y. and Li, L. (2008) Nuclear Dvl, c-Jun, beta-catenin, and TCF form a complex leading to stabilization of beta-catenin-TCF interaction. J. Cell Biol. **180**, 1087–1100.

219 Hsieh, C.-S., Ko, C.-Y., Chen, S.-Y., Liu, T.-M., Wu, J.-S., Hu, C.-H. and Sun, C.-K. (2008) In vivo long-term continuous observation of gene expression in zebrafish embryo nerve systems by using harmonic generation microscopy and morphant technology. J. Biomed. Opt. **13**, 064041.

220 McMahon, C., Gestri, G., Wilson, S. W. and Link, B. A. (2009) Lmx1b is essential for survival of periocular mesenchymal cells and influences Fgf-mediated retinal patterning in zebrafish. Dev. Biol. **332**, 287–298.

221 Roy, S., Wolff, C. and Ingham, P. W. (2001) The u-boot mutation identifies a Hedgehog-regulated myogenic switch for fiber-type diversification in the zebrafish embryo. Genes Dev. **15**, 1563–1576.

222 Teoh, P.-H., Shu-Chien, A. C. and Chan, W.-K. (2010) Pbx1 is essential for growth of zebrafish swim bladder. Dev. Dyn. Off. Publ. Am. Assoc. Anat. **239**, 865–874.

223 Lawrence, C., Adatto, I., Best, J., James, A. and Maloney, K. (2012) Generation time of zebrafish (Danio rerio) and medakas (Oryzias latipes) housed in the same aquaculture facility. Lab Anim. **41**, 158–165.

224 Pugach, E. K., Li, P., White, R. and Zon, L. (2009) Retro-orbital injection in adult zebrafish. J. Vis. Exp. JoVE.

225 Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**, 671–675.

226 Giarratana, M.-C., Rouard, H., Dumont, A., Kiger, L., Safeukui, I., Le Pennec, P.-Y., Francois, S., Trugnan, G., Peyrard, T., Marie, T., et al. (2011) Proof of principle for transfusion of in vitro-generated red blood cells. Blood **118**, 5071–5079.

227 Kearney, J. F., Radbruch, A., Liesegang, B. and Rajewsky, K. (1979) A new mouse myeloma cell line that has lost immunoglobulin expression but permits the construction of antibody-secreting hybrid cell lines. J. Immunol. Baltim. Md 1950 **123**, 1548–1550.

228 Köhler, G. and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature **256**, 495–497.

229 Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal **17**, pp. 10–12.

230 Trapnell, C., Pachter, L. and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. **25**, 1105–1111.

231 Anders, S., Pyl, P. T. and Huber, W. (2014) HTSeq – A Python framework to work with high-throughput sequencing data. bioRxiv 1–5.

232 Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. Genome Biol. **11**, R106.

233 Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. **26**, 139–140.

234 Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., et al. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature **481**, 389–393.

235 ENCODE Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. and Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature **489**, 57–74.

236 Huang, S.-M. A., Mishina, Y. M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G. A., Charlat, O., Wiellette, E., Zhang, Y., Wiessner, S., et al. (2009) Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. Nature **461**, 614–620.

237 Essner, J. J., Breuer, J. J., Essner, R. D., Fahrenkrug, S. C. and Hackett, P. B., Jr. (1997) The zebrafish thyroid hormone receptor alpha 1 is expressed during early embryogenesis and can function in transcriptional repression. Differ. Res. Biol. Divers. **62**, 107–117.

238 Liu, Y. W., Lo, L. J. and Chan, W. K. (2000) Temporal expression and T3 induction of thyroid hormone receptors alpha1 and beta1 during early embryonic and larval development in zebrafish, Danio rerio. Mol. Cell. Endocrinol. **159**, 187–195.

239 Van Beeren, H. C., Jong, W. M. C., Kaptein, E., Visser, T. J., Bakker, O. and Wiersinga, W. M. (2003) Dronerarone acts as a selective inhibitor of 3,5,3'-triiodothyronine binding to thyroid hormone receptor-alpha1: in vitro and in vivo evidence. Endocrinology **144**, 552–558.

240 Baron, M. H., Isern, J. and Fraser, S. T. (2012) The embryonic origins of erythropoiesis in mammals. Blood.

241 Hollyfield, J. G. (1966) Erythrocyte replacement at metamorphosis in the frog, Rana pipiens. J. Morphol. **119**, 1–6.

242 Pelayo, S., Oliveira, E., Thienpont, B., Babin, P. J., Raldúa, D., André, M. and Piña, B. (2012) Triiodothyronine-induced changes in the zebrafish transcriptome during the eleutheroembryonic stage: implications for bisphenol A developmental toxicity. Aquat. Toxicol. Amst. Neth. **110-111**, 114–122.

243 Yamaguchi, M., Takahashi, H. and Wakahara, M. (2000) Erythropoiesis and unexpected expression pattern of globin genes in the salamander Hynobius retardatus. Dev. Genes Evol. **210**, 180–189.

244 Mazurier, C., Douay, L. and Lapillonne, H. (2011) Red blood cells from induced pluripotent stem cells: hurdles and developments. Curr. Opin. Hematol. **18**, 249–253.

245 Kim, M. H., Lin, J., Jhaveri, M. and Koren, A. (2014) Impact of dronedarone treatment on healthcare resource utilization in patients with atrial fibrillation/flutter. Adv. Ther. **31**, 318–332.

246 Crosby, W. H. (1959) Normal functions of the spleen relative to red blood cells: a review. Blood **14**, 399–408.

247 Fischer, U., Ototake, M. and Nakanishi, T. (1998) Life span of circulating blood cells in ginbuna crucian carp (Carassius auratus langsdorfii). Fish Shellfish Immunol. **8**, 339–349.

248 Kuhn, J. M., Rieu, M., Rochette, J., Krishnamoorthy, R., Labie, D., Elion, J., Luton, J. P. and Bricaire, H. (1983) Influence of thyroid status on hemoglobin A2 expression. J. Clin. Endocrinol. Metab. **57**, 344–348.

249 TomHon, C., Zhu, W., Millinoff, D., Hayasaka, K., Slightom, J. L., Goodman, M. and Gumucio, D. L. (1997) Evolution of a fetal expression pattern via cis changes near the gamma globin gene. J. Biol. Chem. **272**, 14062–14066.

250 Mukhi, S., Cai, L. and Brown, D. D. (2010) Gene switching at Xenopus laevis metamorphosis. Dev. Biol. **338**, 117–126.

251 Sassoon, I. and Blanc, V. (2013) Antibody-drug conjugate (ADC) clinical pipeline: a review. Methods Mol. Biol. Clifton NJ **1045**, 1–27.

252 Cho, S. W., Kim, S., Kim, J. M. and Kim, J.-S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. **31**, 230–232.

253 Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science **339**, 819–823.

254 Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J.-R. J. and Joung, J. K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. **31**, 227–229.

255 Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. eLife **2**, e00471.

256 Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. Science **339**, 823–826.

257 Shimizu, N., Kawakami, K. and Ishitani, T. (2012) Visualization and exploration of Tcf/Lef function using a highly responsive Wnt/β-catenin signaling-reporter transgenic zebrafish. Dev. Biol. **370**, 71–85.

258 Meijer, L., Skaltsounis, A.-L., Magiatis, P., Polychronopoulos, P., Knockaert, M., Leost, M., Ryan, X. P., Vonica, C. A., Brivanlou, A., Dajani, R., et al. (2003) GSK-3-selective inhibitors derived from Tyrian purple indirubins. Chem. Biol. **10**, 1255–1266.

259 Kettleborough, R. N. W., Busch-Nentwich, E. M., Harvey, S. A., Dooley, C. M., de Bruijn, E., van Eeden, F., Sealy, I., White, R. J., Herd, C., Nijman, I. J., et al. (2013) A systematic genome-wide analysis of zebrafish protein-coding gene function. Nature **496**, 494–497.

260 De Wit, E. and de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev. **26**, 11–24.

261 Xu, J., Shao, Z., Glass, K., Bauer, D. E., Pinello, L., Van Handel, B., Hou, S., Stamatoyannopoulos, J. A., Mikkola, H. K. A., Yuan, G.-C., et al. (2012) Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell **23**, 796–811.

Appendix

Bcl11a and Sox6 regulate adult and embryonic *globins* in the zebrafish embryo

Bcl11a and Sox6 regulate adult and embryonic *globins* in the zebrafish embryo

A thesis submitted by

Janelle S. Lambert

in partial fulfillment of the requirements
for the degree with Honors of Bachelor of Arts in the field of
Human Developmental and Regenerative Biology

Leonard Zon, MD, Grousbeck Professor of Pediatric Medicine
Harvard Medical School, Children's Hospital Boston

Supervised by Jared Ganis

Harvard University
Cambridge, MA
March 20, 2012

STATEMENT OF RESEARCH

This senior thesis was conducted under the supervision of Jared Ganis and the direction of Leonard Zon, MD at Children's Hospital Boston between January 2011 and March 2012. This thesis is the culmination of one summer of full time work and three semesters of enrollment in SCRB 91r/99.

ACKNOWLEDGEMENTS

My sincerest thanks and gratitude go to my direct supervisor, Jared Ganis, and PI, Leonard Zon, MD, for getting me up to speed and prepared to conduct this thesis in only one year with the Zon laboratory. Their guidance has made this whole process as enjoyable and rewarding as it could possibly be.

I would also like to extend a huge thank you to all of the members of the Zon laboratory who made it such a fun place to be even when I was still there at 9pm at night. You all really know how to welcome a person and help her get over any apprehension about her first lab experience.

Lastly, I appreciate the support I have received from my family and friends, especially my blockmates and teammates, who made the last month struggling to finish bearable.

ABSTRACT

All vertebrates studied to date have multiple *globin* genes that are expressed in different developmental stages. In humans, there is a switch from embryonic to fetal and fetal to adult globin that is largely controlled at the transcriptional level. Understanding hemoglobin switching is of particular interest because of the potential for therapies for hemoglobin disorders such as sickle cell disease and β-thalassemias. This thesis confirms the zebrafish embryonic to larval *globin* switch through *in situ* hybridization and uses the zebrafish as a model organism to evaluate the role of two transcription factors, Bcl11a and Sox6, in the regulation of adult and embryonic *globins* in the zebrafish embryo. These two transcription factors have been implicated in *globin* switching for the adult human and murine cells, but were largely unexplored in the zebrafish embryo before this thesis. In the zebrafish embryo, knockdown and overexpression of both Bcl11a and Sox6 causes decrease and increase, respectively, in both embryonic and adult globin expression. This suggests a different regulatory mechanism than had been previously explored for these transcription factors in adult human and murine cells. Further experiments are necessary to determine if this unknown regulatory mechanism is different due to a divergence in zebrafish regulation versus mammalian or due to a distinct embryonic regulatory mechanism versus an adult one.

TABLE OF CONTENTS

Table of Contents

1. INTRODUCTION

1.1 Hemoglobin switching

During the course of vertebrate development, there are different oxygen demands due to environmental changes in oxygen tension. In order to cope with these changes in demand, all vertebrates studied have evolved different functional hemoglobins with specific oxygen affinities present throughout development. As a result, there are multiple globin genes with developmental stage-specific expression (Brownlie et al., 2003). This complex process of producing distinct globin chains depending on the stage of development is called hemoglobin switching.

Hemoglobin is a tetrameric protein that consists of two α-like and two β-like globin chains. In mammals these α-like and β-like chains are encoded by genes grouped into two separate gene clusters called the α-*globin* locus and the β-*globin* locus. In humans, the α-*globin* locus contains a single embryonic *globin* gene, ζ-*globin*, and two adult α-*globin* genes, while the β-*globin* locus consists of five functional genes (ε, ^{G}γ, ^{A}γ, δ, and β) arranged 5' to 3' in the order of expression during development (McConnell et al., 2011)(Figure 1A). Humans have two developmental *globin* gene switches: the switch from embryonic ε-*globin* to fetal γ-*globin* gene expression, which corresponds with the transition from embryonic (yolk sac) to definitive (fetal liver) hematopoiesis; and then from fetal γ-*globin* to adult β-*globin* gene expression, occurring during the perinatal period (Stamatoyannopoulos, 2005). Similarly, humans also see a switch from embryonic ζ-*globin* gene expression to adult α-*globin* gene expression as the embryo matures (McConnell et al., 2011)(Figure 1B).

Figure 1

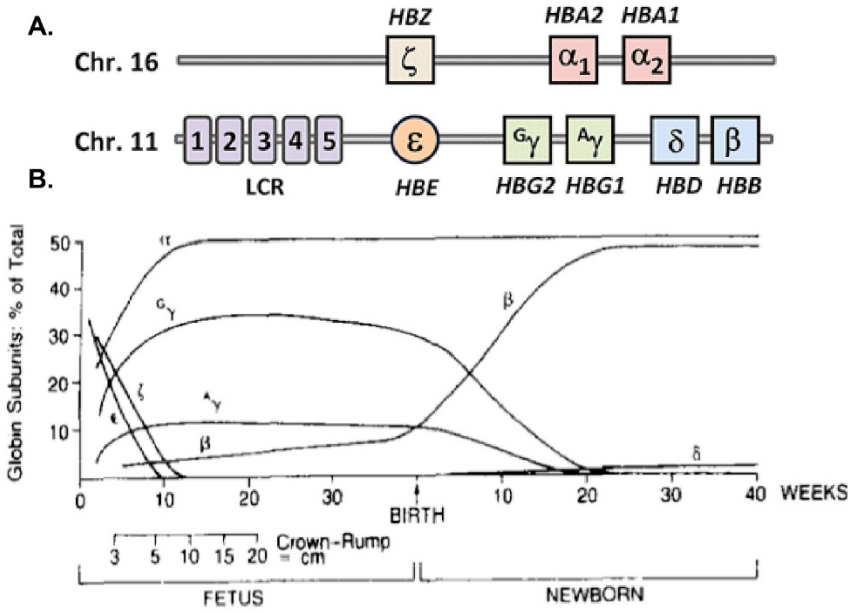


Figure 1. Human *globin* loci and relative expression of globin subunits

(A) Schematic of the α-*globin* locus at chromosome 16 and the β-*globin* locus at chromosome 11. Both loci have genes arranged 5'-3' and in order of expression during development. LCR = locus control region. Modified from Wilber et al., 2011b.
(B) Relative expression level changes of the α- and β-*globin* subunits throughout human development. Reprinted from Palis and Segel, 1998.

1.2 Hemoglobinopathies and their amelioration through persistent fetal hemoglobin expression

The switch from fetal hemoglobin (HbF, $\alpha_2\gamma_2$) to adult hemoglobin (HbA, $\alpha_2\beta_2$) has been of particular interest to researchers because of the implications it carries for those with hemoglobinopathies, diseases in which *globin* genes are mutated or misregulated, such as sickle cell disease and β-thalassemias (Wilber et al., 2011). According to the World Health Organization, more than 200,000 babies are born with sickle cell disease in Africa every year and many of them will die before the age of five years because of anemia and infection by encapsulated organisms, to which they become vulnerable due to abnormal spleen function (Leikin et al., 1989; Orkin and Higgs, 2010). In the United States, an estimated 50,000 people have sickle cell disease (Orkin and Higgs, 2010). Additionally, the thalassemia syndromes are significant causes of mortality and morbidity worldwide, especially in the developing world. Close to two percent of death and disability adjusted life years among children in Asia is due to the thalassemia syndromes (Sankaran et al., 2010).

By the early 1960s, as a result of studying rare patients with both hereditary persistence of fetal hemoglobin (HPFH) and beta-hemoglobinopathies, it was determined that persistent, high-level expression of HbF led to a milder form of these usually severe diseases (Stamatoyannopoulos, 2005). HbF is usually present in low amounts, approximately 1% of all hemoglobin, in all adults, but this concentration of HbF is variable and heritable as a quantitative trait. For those with sickle cell disease, also having HbF concentrations greater than 15% is often associated with relatively benign disease symptoms (Orkin and Higgs, 2010). The normal γ-globin can substitute, without

negative consequences, for the missing β-globin in β-thalassemia and decrease the amount of sickle hemoglobin-associated polymerization in sickle cell disease (Bauer and Orkin, 2011). The clinical benefit of increased levels of HbF for sickle cell disease and β-thalassemia patients has been the driving force for studying hemoglobin switching (Stamatoyannopoulos, 2005).

1.3 Animal models used to study hemoglobin switching

1.3.1 The mouse as a model organism to study hemoglobin switching

Animal models, such as the mouse and, shown here, the zebrafish, are currently being used to study hemoglobin switching. The murine model has been useful with an α-*globin* locus similar to the human with one embryonic ζ-*globin* gene and two adult α-*globin* genes. Slightly different from the human, the murine β-*globin* locus has four genes (εY, βh1, β1, and β2). Similar to humans, the genes on both *globin* loci are arranged generally in their order of expression. Unlike humans, however, mice only have a one developmental *globin* gene switch from embryonic to adult *globin* genes, since they do not have distinct fetal and adult type hemoglobins. Instead, mice express their adult β1- and β2-*globin* genes during both fetal liver and adult bone marrow definitive erythropoiesis (Bauer and Orkin, 2011; McConnell et al., 2011). Although this difference exists, mice, particularly human β-*globin* locus transgenic mice, still provide a valid system for evaluating human developmental *globin* gene regulation. These transgenic mice allow for the study of the regulation of human *cis*-elements in a mouse *trans*-acting environment (Sankaran et al., 2009). Unfortunately, the need to produce transgenic mice

to carry out *in vivo* experiments, as well as produce fetal stage hemoglobin, makes the use of the mouse a challenging and time-consuming process (McConnell et al., 2011).

1.3.2 The zebrafish as a model organism to study hemoglobin switching

Unlike mammals, the zebrafish *globin* genes are located in two loci, major and minor, and are arranged in α/β pairs oriented in a 5'-3' to 3'-5' manner (Figure 2A). Similar to humans, however, Ganis *et al*. recently determined, through the use of quantitative real-time PCR (qPCR) data on the relative levels of zebrafish α- and β-globin expression through development, that zebrafish have two developmental *globin* switches: one from embryonic to larval and another from larval to adult globins (Ganis et al., unpublished)(Figures 2B and 2C). Additionally, as will be discussed further in this thesis, I completed *in situ* hybridizations (ISHs) through 5 days post fertilization (dpf), providing temporal and spatial information about the globin expression patterns, confirming the qPCR data and providing further evidence for the existence of a zebrafish embryonic to larval globin switch.

The zebrafish also proves to be a relevant model organism for the study of hematopoiesis as the developmental processes and genetic programs of hematopoiesis are highly conserved in zebrafish, using many of the same regulatory regions and genes seen in humans such as *gata1* and *runx1* (Hsia and Zon, 2005). Additionally, zebrafish start producing hemoglobin at an early age with erythrogenesis beginning at approximately 15 hours post fertilization (hpf) and blood cells containing hemoglobin are found in circulation at 24 hpf. This early production, however, does not indicate necessity, as zebrafish embryos can survive without hemoglobin or red blood cells through at least 15

dpf by passive diffusion of oxygen (Rombough and Drader, 2009). This feature of zebrafish hematopoetic development allows for the identification of mutations and detailed loss-of-function experiments that would be otherwise impossible (Paik and Zon, 2010; Ganis et al., unpublished). Also, since the zebrafish has many appealing features such as high fecundity, external fertilization, and transparent embryos, it proves to be an extremely useful organism for *in vivo* imaging of both normal and unusual hematopoiesis (Jing and Zon, 2011).

Figure 2

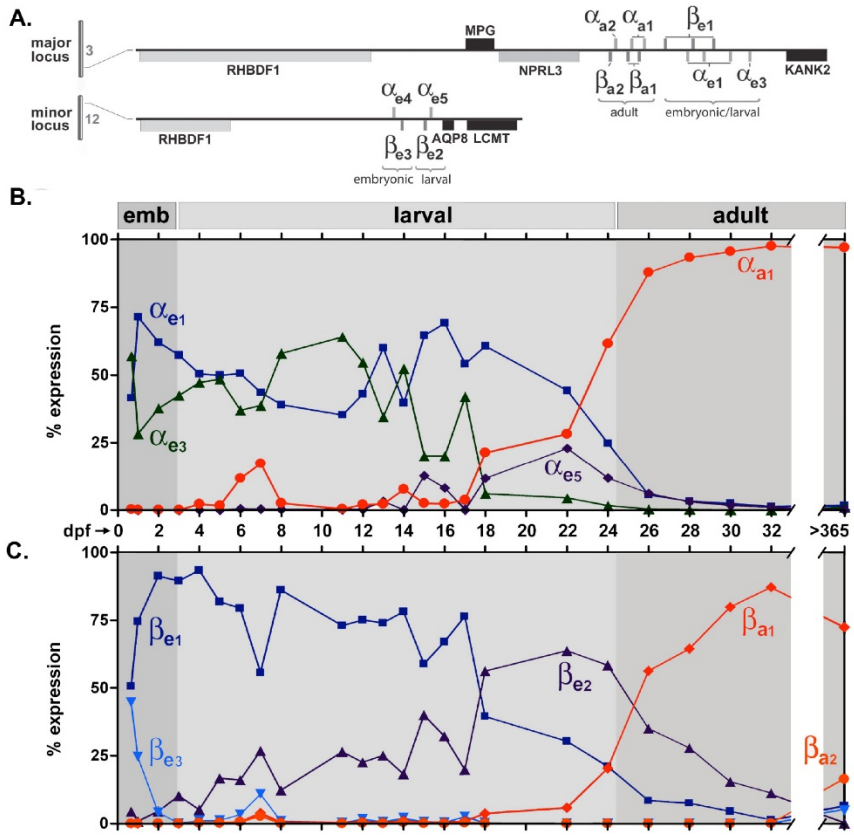


Figure 2. Zebrafish *globin* loci and relative *globin* expression levels throughout development
(A) Major and minor zebrafish *globin* loci located on chromosomes 3 and 12 respectively. The *globin* genes are arranged in α/β pairs and oriented in a 5'-3' to 3'-5' manner. A bidirectional promoter is found between each α/β pair.
Relative expression level changes of the (B) α-*globin* and (C) β-*globin* genes. The stages of *globin* expression are designated by various shades of gray.
Modified from Ganis et al., unpublished.

1.4 Potential mechanisms of hemoglobin switching

The clinical importance of this switch has led to the use of humans cells along with these animal models to provide insight into the mechanisms of hemoglobin switching and HbF regulation. One of the early discoveries resulting from these efforts to determine the control of HbF production were DNA enhancer regions that were initially identified as DNase I hypersensitive sites found upstream of the *globin* genes. These elements, collectively called the locus control region (LCR) were shown to be necessary for high-level *globin* gene expression. Additionally, in transgenic mice, LCR elements were sufficient to obtain a significant level of copy number-dependent β-*globin* gene expression, while deletion caused the silencing of transcription of the *globin* genes (Wilber et al., 2011). In humans, deletion of the LCR results in β-thalassemia distinguished by the lack of β-*gene* expression (Stamatoyannopoulos, 2005).

The discovery of the LCR led to various proposed models of hemoglobin switching, which included chromosomal looping, tracking, and linking models (Stamatoyannopoulos, 2005). Chromosomal looping suggests that the LCR folds to form an active site that binds transcription factors. This structure then "loops," bringing the LCR in close proximity to the target gene. This allows for interaction of the transcription factors bound to the LCR with the promoter of the gene (Palstra et al., 2008; Stamatoyannopoulos, 2005). The tracking model suggests that the LCR acts as a loading platform for a DNA-tracking protein that will then move along the chromatin toward the promoter (Palstra et al., 2008). The linking model is similar to the tracking model in that the LCR acts as a loading platform for a DNA-binding protein. Instead of moving along the chromatin, however, the protein facilitates the binding of additional proteins in the

direction of the promoter, covering the chromatin (Palstra et al., 2008). Regardless of intensive molecular studies, no one model has been agreed upon, and multiple models may cooperate to compose the true mechanism (Wilber et al., 2011). Additionally, the proteins modulating the switch and their interactions are not entirely understood (Xu et al., 2010).

1.5 Molecular control of the fetal to adult hemoglobin switch

1.5.1 BCL11A

There has been substantial evidence indicating that the fetal to adult hemoglobin switch is influenced by multiple *cis*-acting elements as well as many transcription factors acting as multiprotein complexes (Wilber et al., 2011). BCL11A is one transcription factor that has been suggested in multiple studies to have an important role in switching for the β-*globin* gene locus (Sankaran et al., 2008; Sankaran et al., 2009; Wilber et al., 2011; Xu et al., 2010; Zhou et al., 2010). BCL11A, a zinc-finger transcription factor, was initially only associated with a role for B-lymphocyte production. Upon the completion of genome-wide association studies to find common genetic variants associated with differences in HbF levels, a region within the *BCL11A* gene was one of three major genomic loci implicated (Sankaran, 2011). BCL11A was eventually shown to act as a repressor of the γ-*globin* gene. This was demonstrated using shRNA-mediated knockdown of BCL11A, which resulted in the increase of HbF in primary adult human erythroid cells (Sankaran et al., 2008). Similar results were demonstrated in the mouse as knockdown of BCL11A caused an increase in the endogenous embryonic hemoglobin in adult murine cells (Sankaran et al., 2009). Sankaran *et al.* also determined that BCL11A

protein levels were related to the developmental stage of expression. Primitive and fetal liver erythroid cells expressing high levels of γ-*globin* had little or no expression of the full-length isoforms of BCL11A (-L and -XL), expressing the shorter variants (-XS and -S) instead. Abundant expression of the full-length forms was found in the adult erythroid cells (Liu et al., 2006; Sankaran et al., 2008; Sankaran et al., 2009). At the protein level the L and XL isoforms are indistinguishable because of their similar size, however, at the mRNA level, the XL isoform is seven times more abundant than the L isoform in adult human erythroid progenitors (Sankaran et al., 2008; Xu et al., 2010). This suggests that the XL isoform is predominant (Xu et al., 2010), and may be more effective in its role in the cell. Additionally, BCL11A associates directly with chromatin at the human β-*globin* locus in primary adult human cells as well as interacts with the Mi-2/NuRD chromatin remodeling complex and the transcription factors GATA1 and FOG1 (Xu et al., 2010).

1.5.2 SOX6

Another transcription factor thought to be critical in regulating the fetal to adult *globin* switch is SOX6. SOX6, a member of the Sry-related high-mobility group (HMG) box transcription factors, was first linked to mouse embryonic *globin* gene expression through analysis of the *Sox6*-deficient mouse. While SOX6 normally represses the εγ-*globin* gene in definitive erythropoiesis, in *Sox6*-deficient mice εγ-*globin* gene was continually expressed until birth (Yi et al., 2006). Additionally, transplantation of fetal liver cells from *Sox6*-deficient mice into wildtype adult mice showed that SOX6 represses εγ-*globin* and, to a lesser extent, βh1-*globin* expression (Cohen-Barak et al., 2007). Recent evidence from Xu *et al.* suggests that BCL11A-mediated silencing occurs through

cooperation with SOX6 as they are coexpressed and physically interact with each other. Additionally, knockdown of both BCL11A and SOX6 in adult primary human erythroblasts leads to greater HbF production than seen with either knockdown alone (Xu et al., 2010).

1.6 Experimental Approach

While the roles of BCL11A and SOX6 have been investigated in adult cells, the roles of BCL11A and SOX6 as regulators of *globin* gene expression in the embryo remains unexplored. Preliminary data from Jared Ganis in the Zon laboratory shows that a small number of adult *globin* expressing cells are present in the embryo. Additionally, Bcl11a is present in the embryo and knockdown in the embryo results in a decrease in adult *globin* expression, while overexpression in the embryo results in an increase in adult *globin* expression (Ganis, unpublished data). This suggests that Bcl11a has a regulatory role for both adult and embyronic *globin* expression. Through microinjections of cDNA and antisense morpholino (MO) oligonucleotides I completed overexpression and knockdown of Bcl11a and Sox6 in the zebrafish embryo demonstrating a regulatory role for both Bcl11a and Sox6 for adult and embryonic *globin* expression in the embryo.

2. METHODS AND MATERIALS

2.1 Zebrafish maintenance and staging

Zebrafish were raised and maintained under approved laboratory conditions. For all experiments Tübingen zebrafish were used. Zebrafish were staged as described for every developmental time point of interest (Kimmel et al., 1995). At developmental time points prior to hatching from the chorions, the embryos were subjected to a pronase solution for seven minutes to remove the chorions. The embryos were then washed three times with E3 fish water before being fixed in 4% paraformaldehyde (PFA) at 4°C overnight. Following fixation, the embryos were washed three times in methanol for dehydration and stored at -20°C until *in situ* hybridization.

2.2 *In situ* hybridization

Whole-mount ISH was performed, with some modification, according to a three-day protocol (Thisse and Thisse, 2008) manually or by a BioLane robot. This procedure allowed for the high-resolution visualization, at the single-cell level, of specific genes in their site of expression during development. This was accomplished through the hybridization of digoxigenin-labeled antisense RNA probes to complementary mRNA in the embryo.

On Day 1, embryos were rehydrated from their storage in methanol through a series of washes of methanol/PBT solutions at 75%/25%, 50%/50%, and 25%/75% followed by four five-minute PBT washes. All embryos fixed at 24 hpf or later were then incubated in a bleach solution (1% H_2O_2, 0.8% KOH, 0.1% Tween and ddH_2O) to

remove the pigmentation in the embryos that would prevent accurate visualization of the expression pattern. After all of the pigmentation was removed in the embryos 24 hpf or later, all embryos, regardless of age, were digested with protinease K at 10μg/mL from two to twenty minutes depending on the developmental stage of the embryo; older embryos require a longer digestion period. This step allowed for the probe to penetrate the tissues of the embryos. This protinease K digestion was stopped with three quick washes of PBT and a 20-minute incubation with a PFA/glutaraldehyde solution. After four five-minute washes with PBT to remove the excess PFA, the embryos were prehybridized for between 45 minutes and 4 hours at 70°C with a hybridization buffer (50% formamide, 5x SSC, 50 μg/mL tRNA, 50 μg/mL heparin, 0.1% Tween and ddH_2O). Following prehybridization, the solution was replaced with hybridization buffer containing the addition of a specific antisense digoxigenin-labeled RNA probe for the globin of interest. These embryos were left to hybridize overnight at 70°C.

On Day 2, embryos were gradually changed from the hybridization buffer to 2x SSC in a series of ten minute, 70°C hybridization buffer/2x SSC washes at 75%/25%, 50%/50%, 25%/75%, and 0%/100%. The hybridization buffer in these washes was without tRNA, heparin or probe. Two 30-minute washes in 0.2x SSC at 70°C were then followed by another series of washes at room temperature replacing 0.2x SSC with PBT gradually through 0.2x SSC/PBT solutions at 75%/25%, 50%/50%, 25%/75%, and 0/100%. After these washes the embryos were incubated at room temperature for at least an hour in blocking solution (2% lamb serum, 2 mg/mL of bovine serum albumin, and PBT). The embryos were then stored overnight at 4°C in an antibody solution which was composed of a 1:5000 dilution of anti-digoxigenin antibody in blocking solution.

On Day 3, the antibody solution was removed and the embryos washed five times, 15 minutes per wash at room temperature, in PBT. After this the embryos underwent three 5-minute washes with staining buffer (100mM Tris pH 9.5, 50mM $MgCl_2$, 100mM NaCl, 0.1% Tween, and ddH_2O). Digoxigenin staining followed through incubation in staining solution (staining buffer with 0.175 mg/mL BCIP and 0.225 mg/mL NBT). The embryos remained in this staining solution until the proper staining intensity was achieved; this was anywhere between 20 minutes and 3.5 hours. The staining reaction was stopped by three quick washes in PBT and two in stop solution (PBT pH 4.5). These embryos were then stored in stop solution at 4°C until pictures were taken.

2.3 Morpholinos (knockdown)

MOs were obtained through Gene Tools, LLC. MOs are chemically modified antisense oligonucleotides that work through an RNase-H independent process that either blocks translational initiation by binding to sequences flanking and including the ATG start site (an ATG MO) or prevents accurate pre-mRNA processing by blocking exon and intron boundaries from splicing mechanisms (an exon MO) (Nasevicius and Ekker, 2000). For optimal binding, MOs were designed to be 24-25 nucleotides long, containing approximately 50 percent G/C content with no internal hairpins. Additionally, each sequence was tested through BLAST to ensure specific binding only to the target mRNA. Bcl11aa and Bcl11ab MOs were acquired from Jared Ganis, while I designed the Sox6 MO. The sequences were as follows (predicted start codons underlined if an ATG MO): Bcl11aa Exon 1 MO (targeting the exon-intron boundary of Exon 1 in Bcl11a), 5'-TGTCTTTACTTACGCGAAAAATCC-3'; Bcl11ab ATG MO, 5'-

CTCACATTGGGAAGCGCCAGAGTC-3'; Sox6 ATG MO, 5'-TGGCTTGCTTGGAAGACATGATTCT-3'

2.4 Expression vectors (overexpression)

Four overexpression vectors were generated, each driving expression of a different transcription factor or variant: zebrafish *bcl11a*, human *BCL11A-L*, human *BCL11A-XL*, and human *SOX6*. Human genes were used in overexpression experiments to show conservation between zebrafish and human function of these transcription factors.

BCL11A-L, *BCL11A-XL*, and *SOX6* cloned into pEF1/V5-His were obtained from Jian Xu in the Orkin laboratory. All three of these genes were cloned into the Gateway entry pENTR 5'-TOPO vector (Invitrogen) through PCR amplification. *bcl11a* already cloned into the Gateway entry pENTR 5'-TOPO was acquired from Jared Ganis in the Zon laboratory. After purification through Mini-Prep (Qiagen), correct incorporation into the entry vector was confirmed via sequencing. Each of these entry vectors were combined with the 5' vector containing *drl3*, vector #302, and pDest-drlGFP (Kwan et al., 2007) to assemble the final expression vector using LR Clonase II Plus (Invitrogen). The 5' vector containing *drl3,* a fragment of the *draculin (drl)* promoter, was cloned by Christian Mosimann in the Zon laboratory. The pDest-drlGFP vector was created by Jared Ganis using vector #394 from the Tol2kit (Kwan et al., 2007) and the *drl* promoter cloned by Christian Mosimann. *drl* is an early marker of the intermediate cell mass (ICM), with expression beginning at the late blastula stage and continuing through 48 hpf

(Hsia and Zon, 2005). As the embryos in this thesis were staged at 24 hpf, the *drl* promoter is useful tool to drive overexpression and GFP expression.

2.5 Microinjections

For all injections, embryos were fixed at 24 hpf. I tested injections at three developmental time points: 16 ss, 25 ss, and 24 hpf. I determined, by looking for severe morphological changes as well as considering the difficulty in identifying that there was a change in stain intensity, that 24 hpf was the best time point for all future experiments.

2.5.1 Knockdown

Embryos were injected, with one of the three MOs from 2.3, into the yolk at the one-cell stage (within approximately 45 minutes post fertilization). For each MO a range of concentrations were injected to identify a dose that maximizes the phenotype without inducing severe morphological defects. For the ATG MOs, Bcl11ab and Sox6, concentrations of 2ng, 4ng, and 6ng were used. Generally, 4ng was the most effective without causing severe morphological defects. 6ng injections of Sox6 MO caused death in all embryos injected. For the Bcl11a Exon 1 MO, concentrations of 4ng, 8ng, and 12ng we attempted. 8ng was the most effective dose. All injected embryos, as well as uninjected siblings as a control, were fixed at 24 hpf for ISH.

2.5.2 Overexpression

Embryos were injected, with one of the four expression vectors constructed as discussed in 2.4, into the cell at the one-cell stage. The vectors were injected at 25 ng/μL

in combination with Tol2 mRNA (15 ng/μL). The injected embryos were observed at 24 hpf using a fluorescent microscope. These embryos were separated according to high, low, or no GFP expression as indicated by Figure 3. The embryos with no GFP expression were discarded, while those with high and low expression were fixed at 24 hpf for ISH.

Figure 3

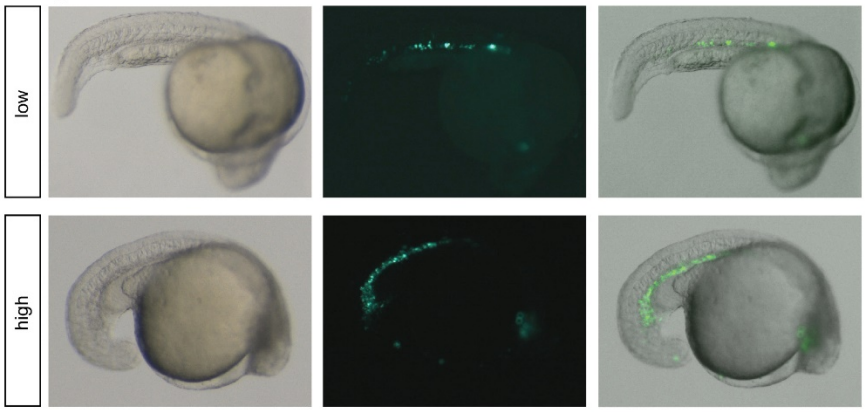


Figure 3. Low and high GFP expression in transiently injected zebrafish
Embryos injected with the *BCL11A-XL* expression vector displaying fluorescent staining representative of the two groups utilized in scoring overexpression: high and low expression. Similar expression patterns are seen in *bcl11a, BCL11A-L,* and *SOX6* overexpression.

2.6 Statistical Analysis

Fisher's exact test (Stata/SE 10.0, Microsoft Excel) was done to determine if the increase or decrease in expression seen between uninjected and injected sibling embryos stained for the same globin was significant. Since some sample sizes were below five, Fisher's exact text was used instead of a chi-squared test. The increase or decrease in *globin* expression was considered significant for values $p<0.05$

3. RESULTS

3.1 Zebrafish have an embryonic to larval *globin* switch

qPCR data from the Zon laboratory, on the relative levels of zebrafish α- and β-*globin* expression through development, indicates that hemoglobin switching occurs in two stages: embryonic to larval and larval to adult (Ganis et al., unpublished) (Figures 2B and 2C). While this qPCR provides quantitative information on the *globin* expression and allows for a long-term overview of the changes in expression levels, information concerning the expression patterns of the globins was missing. To confirm the qPCR data collected as well as provide further information about the spatial and temporal expression patterns of the *globins* at specific developmental time points, I completed ISHs beginning at the 16 somite stage (ss) through 5 dpf for the nine unique *globins*: six embryonic (α_{e1}, α_{e3}, α_{e5}, β_{e1}, β_{e2}, and β_{e3}) and three adult (α_{a1}, β_{a1}, and β_{a2}) (Figure 4).

These ISHs confirm that there is a swtich from embryonic to larval *globins* in the zebrafish. One of the features of this switch is demonstrated in the ISHs by the decreasing intensity of the β_{e3} staining, particularly between 48 hpf and 3 dpf where it shifts from some stain to no stain within 24 hours (Figure 4). This dramatic decrease in β_{e3} is also shown in the qPCR data (Figure 2C). The ISHs also indicate a striking increase in β_{e2}. From the first 16 ss image through 5 dpf, there is a gradual increase in the intensity of the stain. Again, this data is supported by the qPCR data. All of the adult *globins* show little to no staining at the time points for which ISHs were done. This data is unsurprising as the qPCR data indicates that adult *globins* do not start to be strongly expressed until around 22 dpf. A small amount of staining was expected for the adult *globins* since there

are a few adult *globin*-expressing cells in the embryo (Ganis, unpublished data). At 24 hpf, for example, both α_{a1} and β_{a1} have a few cells stained indicating that they are expressing the respective adult *globin* (Figure 4). The sensitivity that ISH provides to allow these adult *globin*-expressing cells to be stained is one of the advantages of ISH over qPCR. Such a low number of cells would not be detected by qPCR or the data would be considered unreliable.

Figure 4

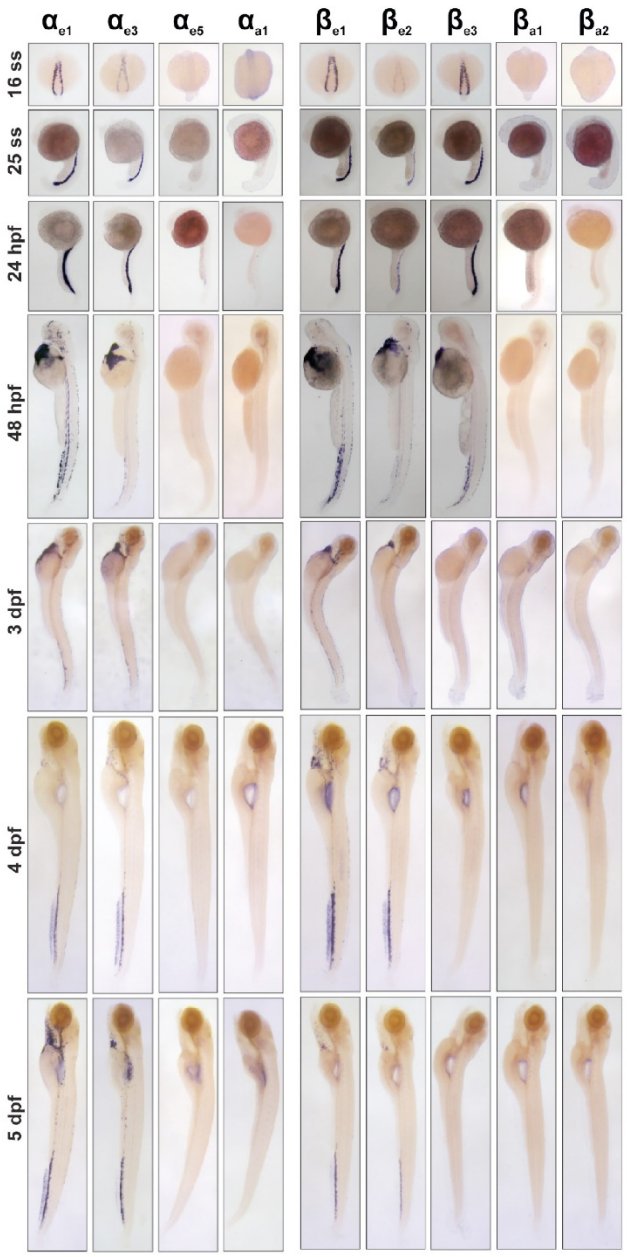


Figure 4: *globin* expression by ISH through 5 dpf
Staining for nine unique *globins* – 4 α-*globins* and 5 β-*globins* – at multiple time points from 16 ss through 5 dpf. These images visualize the embryonic to larval *globin* switch, one of the two *globin* switches seen in the zebrafish.

3.2 Bcl11a regulates *globin* expression in the zebrafish embryo

3.2.1 Regulation of adult *globin* expression in the zebrafish embryo

The ISH data reported in 3.1, as well as unpublished data from Jared Ganis in the Zon laboratory, demonstrates that there are a small number of adult *globin* expressing cells in the embryo (Figure 4). Additionally, knockdown of BCL11A in the zebrafish embryo causes a decrease in α_{a1} expression, while overexpression causes an increase (Ganis, unpublished data). This, along with the previously discussed studies showing that BCL11A knockdown in adult human and murine cells causes an increase in embryonic *globins* (Sankaran et al., 2008; Sankaran et al., 2009), suggests that Bcl11a may have a regulatory role for both embryonic and adult *globins* in the zebrafish embryo. I confirmed the data from Jared Ganis by injecting wildtype zebrafish embryos at the one-cell stage with a Bcl11a MO and completing ISHs for 24 hpf.

Uninjected sibling embryos were used as controls. These embryos had approximately four to six α_{a1} expressing cells located in the ICM, where hematopoiesis occurs at 24 hpf (Orkin and Zon, 2008)(Figure 5). The embryos injected with the Bcl11a MO demonstrated a complete lack of staining in the ICM region in the majority of the injected embryos (13 of 19 embryos, Fisher's exact p=0.00036), indicating that there were no α_{a1} expressing cells present. This suggests that a knock down of Bcl11a led to a decrease in α_{a1} expressing cells.

To further elucidate the effects of BCL11A on α_{a1} expression in the zebrafish embryo, I created an expression vector to over express zebrafish *bcl11a* in all red blood cells as well as express GFP. I injected this vector into the cell of the embryo at the one-cell stage and at 24 hpf selected those embryos expressing GFP for further experiments.

Results

Uninjected embryos generally had 4-6 α_{a1} expressing cells, while 80 percent of those injected with the expression vector saw an increase in α_{a1} expressing cells (12 of 15 embryos, Fisher's exact p=1.16 x 10^{-7}); approximately 12 α_{a1} expressing cells was typical of injected embryos (Figure 5). This data demonstrates that overexpression of *bcl11a* causes an increase in α_{a1} expressing cells.

In order to confirm the conservation of function between the zebrafish and human proteins, I repeated the overexpression experiments with two additional vectors, each containing one of the two major isoforms that BCL11A is expressed as in human adult cells: BCL11A-L and BCL11A-XL (Sankaran et al., 2008). For both of these isoforms, there was an increase in α_{a1} staining in the injected embryos as compared to the uninjected embryos from the same clutch. Approximately two-thirds of the BCL11A-L injected embryos had an increase in α_{a1} expressing cells (23 out of 34 embryos, Fisher's exact p = 6.80 x 10^{-14}) (Figure 5). Likewise, 60 percent of the BCL11A-XL injected embryos demonstrated an increase in α_{a1} staining (25 out of 40 embryos, Fisher's exact p = 4.05 x 10^{-15}) (Figure 5). This similarity in the phenotypes shown by the zebrafish and human BCL11A proteins suggest that they act in the same manner in the zebrafish embryo, indicating function is conserved between the human and zebrafish BCL11A proteins.

In order to try to more fully understand the role of BCL11A for the adult globins, I attempted to consider the effect of BCL11A overexpression on β_{a1} expressing cells. Only one out of the five injected embryos demonstrated an increase in β_{a1} staining, indicating that there was no significant change in the number of β_{a1} expressing cells (Fisher's exact p = 1).

Figure 5

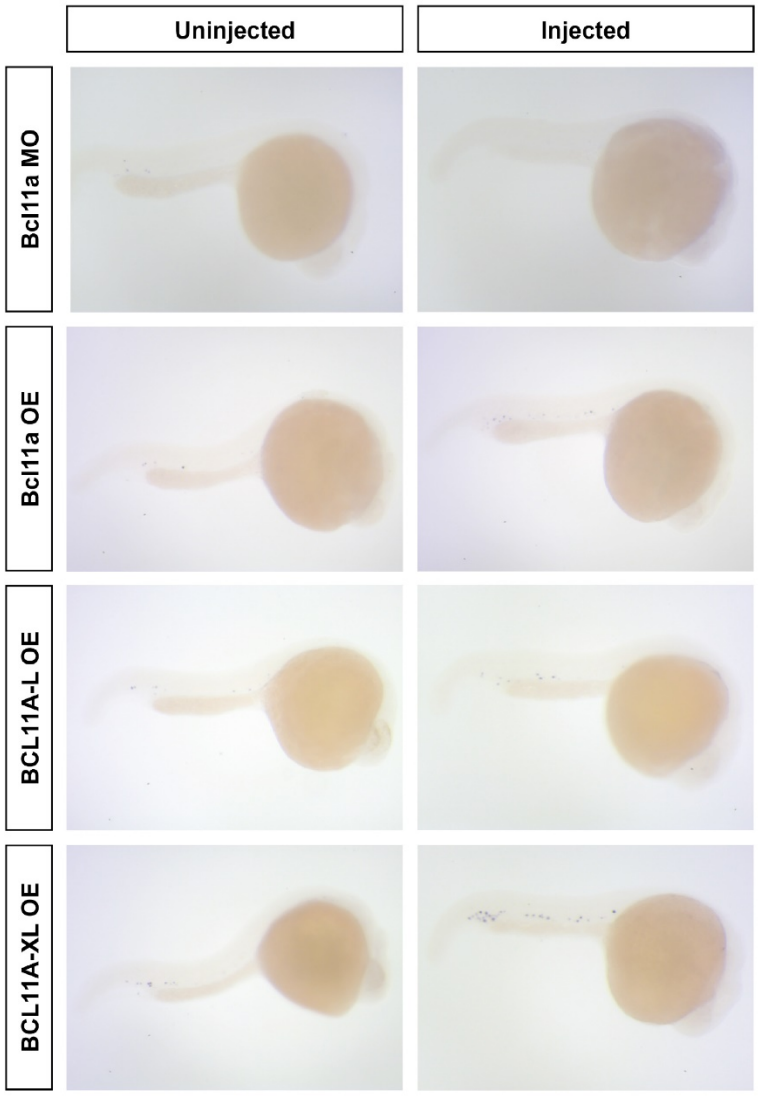


Figure 5. Changes in α_{a1} expression due to Bcl11a knockdown and overexpression
Knockdown (MO) in Bcl11a caused a decrease in α_{a1} expressing cells, while overexpression (OE) led to an increase in α_{a1}. Functional conservation was observed between the zebrafish and human forms of Bcl11a.

3.2.2 Regulation of embryonic *globins* in the zebrafish embryo

With data suggesting that Bcl11a has a regulatory role in the zebrafish embryo for α_{a1}, I decided to explore the possible regulatory role of Bcl11a for embryonic α and β globins. I attempted to complete the knockdown and overexpression injections and ISHs for globins representative of all of the globin expression patterns generally seen in the wildtype zebrafish. β_{e2} was the best embryonic/larval globin with which to examine staining intensity, since its wildtype stain has a moderate intensity that makes it easy to see both increases and decreases in globin expression patterns.

To examine the role of Bcl11a concerning the embryonic globins in the zebrafish, I did overexpression and knockdown for Bcl11a and looked at the α_{e3} and β_{e2} globin expression patterns. With overexpression of zebrafish *bcl11a*, a little less than half of the injected embryos demonstrated an increased intensity of staining, indicating that there were more α_{e3} expressing cells (7 of 15 embryos, Fisher's exact $p = 0.00096$) (Figure 6). This increased staining is characterized by a more densely stained ICM, particularly in the posterior region of the embryo. With the difficulty of identifying this slight difference in stain, it was necessary to be very conservative in determining which embryos showed an increase in α_{e3} expressing cells.

Overexpression of the zebrafish *bcl11a* caused an increase in β_{e2} staining for more than two-thirds of the injected embryos (21 out of 30 embryos, Fisher's exact $p = 4.88 \times 10^{-6}$) (Figure 6). Likewise, overexpression of both human *BCL11A* isoforms lead to an increase in β_{e2} expressing cells. Over 70 percent of the *BCL11A-L* injected embryos saw an increase (46 out of 63 embryos, Fisher's exact $p = 3.70 \times 10^{-15}$), while nearly 90 percent of the *BCL11A-XL* injected embryos did (71 out of 81 embryos, Fisher's exact *p*

= 1.99 x 10^{-32}) (Figures 6). The increased β_{e2} expression of the human *BCL11A-XL* overexpression was particularly striking because nearly every injected embryo displayed intensely darker staining. This may hint at BCL11A-XL being more effective than BCL11A-L, which is suggested by the overabundance of the XL form as compared to the L form in the human adult cell (Xu et al., 2010).

Knockdown of Bcl11a led to decreased staining in every embryo indicating that there was a decrease in β_{e2} expressing cells (4 out of 4, Fisher's exact p = 0.00138) (Figure 6). This data, together with the overexpression data suggests that Bcl11a has a regulatory role in the zebrafish embryo for both the embryonic and adult globins.

Figure 6

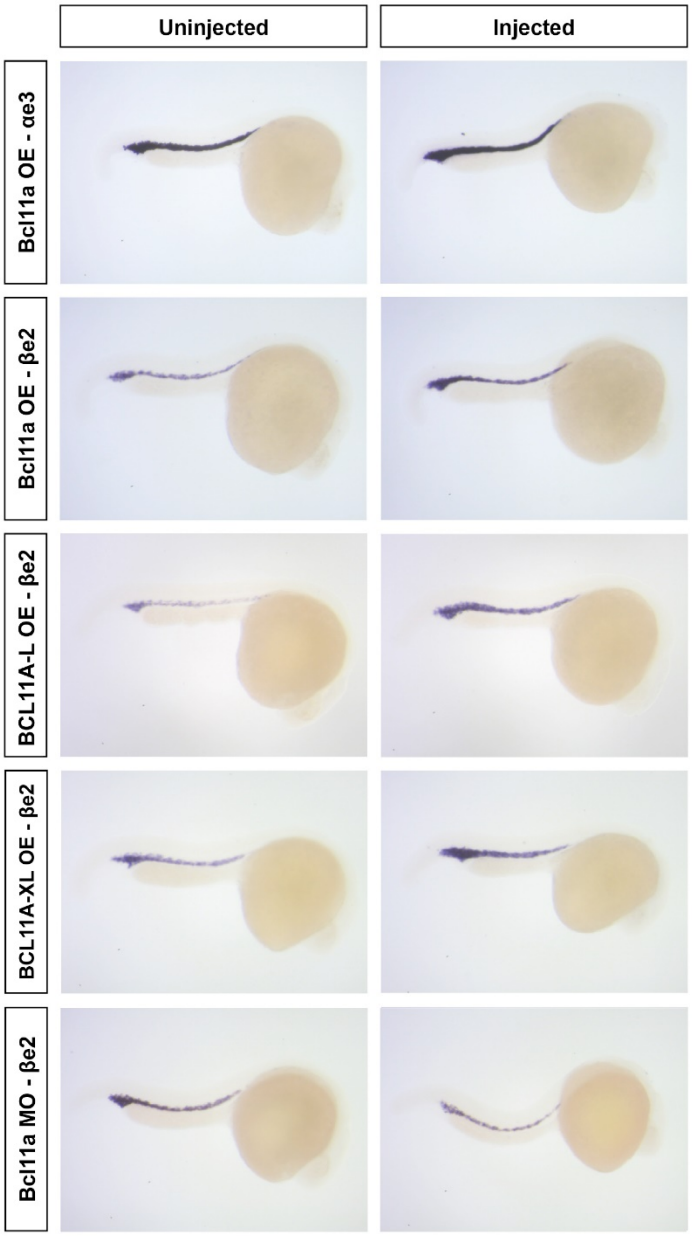


Figure 6. Changes in embryonic *globin* expression due to Bcl11a overexpression and knockdown. An increase in embryonic *globins* is seen upon *Bcl11a* overexpression while a decrease in expression is displayed with Bcl11a knockdown. β_{e2} proves to be the most effective globin to utilize to examine changes in expression. Overexpression = OE. Knockdown = MO.

3.3 Sox6 regulates *globins* in the zebrafish embryo

3.3.1 Regulation of *adult* globins in the zebrafish embryo

Since it has been suggested that BCL11A activity is mediated through SOX6(Xu et al., 2010), I conducted knockdown and overexpression of Sox6 in the zebrafish embryo to explore its possible regulatory role.

Overexpression of human *SOX6* in the zebrafish embryo resulted in an increase in α_{a1} expressing cells in a little over half of the injected embryos (16 of 30 embryos, Fisher's exact $p = 0.00019$) (Figure 7). While uninjected embryos showed an average of about 4 α_{a1} expressing cells, the injected embryos had approximately 8 α_{a1} expressing cells in those with increased staining. I attempted to do knockdown of Sox6 in the zebrafish embryo, but not enough embryos were remaining following staging, fixation, and ISH to determine the effect, if any, of Sox6 knockdown in the embryo on the number of α_{a1} expressing cells. Additionally, similar to the result seen in the Bcl11a overexpression examining the effect on β_{a1} globin expression, with *SOX6* overexpression, less than half of the embryos showed a relatively large number of β_{a1} expressing cells (5 out of 11 embryos). When looking at the uninjected embryos, however, there was a similar ratio of embryos with a large number of β_{a1} expressing cells versus those with a smaller number, indicating that there was no significant change in the β_{a1} staining (Fisher's exact $p = 1$). While this is a limited amount of data, it still suggests that SOX6 has a regulatory role in the zebrafish embryo for at least α_{a1}.

3.3.2 Regulation of embryonic *globins* in the zebrafish embryo

Results

To explore the role of SOX6 in the regulation of embryonic globins in the zebrafish embryo, I did knockdown and overexpression of Sox6, then completed ISHs for α_{e3}, β_{e2} and β_{e3}. Knockdown of Sox6 led to decreases in staining in the ICM for both α_{e3} and β_{e3}. 80 percent of the injected embryos stained for α_{e3} expressing cells saw a decrease (8 out of 10 embryos, Fisher's exact p =0.00073), while just over 60 percent of the injected embryos stained for β_{e3} also saw a decrease (8 out of 13 embryos, Fisher's exact p = 0.0016) (Figures 7). The decrease in expressing cells for both α_{e3} and β_{e3} was characterized by thinner and less dense staining. This data suggests that knockdown of Sox6 causes a decrease in the embryonic globins.

Overexpression of human *SOX6* led to increases in the intensity of staining for α_{e3}, β_{e2} and β_{e3}. An increase in the intensity of staining, particularly noticeable at the most posterior section of the ICM, was seen in over 65 percent of the injected embryos stained for α_{e3} expressing cells (36 out of 55, Fisher's exact $p = 6.29 \times 10^{-12}$) and in a little less than 65 percent of the injected embryos stained for β_{e3} (29 out of 45, Fisher's exact $p = 8.78 \times 10^{-10}$) (Figure 7). Since this staining pattern is not easily noticed and looks somewhat like overstaining, to be sure of the increase in staining seen, I repeated the overexpression for β_{e2}. Nearly 80 percent of the injected embryos demonstrated an increase in β_{e2} expressing cells (26 out of 33, Fisher's exact $p = 3.83 \times 10^{-13}$) (Figure 7). This data, together with the Sox6 data on the adult globins, suggests that Sox6 has a regulatory role for both embryonic and adult globins in the zebrafish embryo.

Figure 7

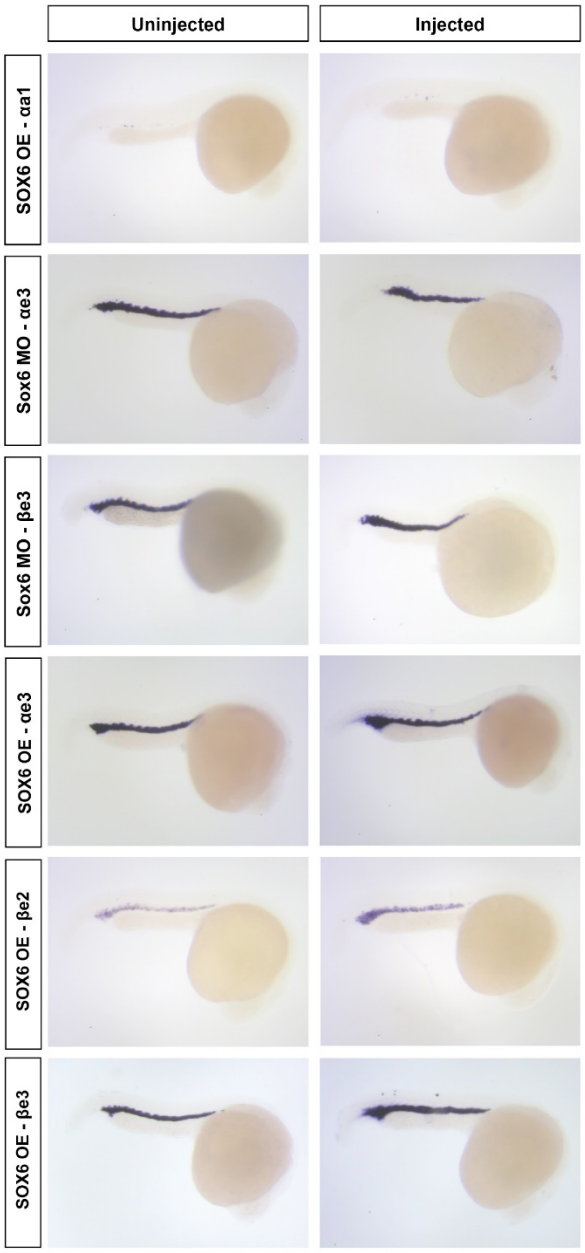


Figure 7: Sox6 regulates adult and embryonic *globin* expression in the zebrafish embryo
Overexpression of *SOX6* caused an increase in α_{a1} expressing cells as well as embryonic *globins*. Knockdown of Sox6 led to a decrease in embryonic *globins*. Conservation of function is seen for Sox6 between humans and zebrafish as indicated by increased expression for both adult and embryonic *globins* when human *SOX6* is overexpressed. Overexpression = OE. Knockdown = MO.

4. DISCUSSION

4.1 Zebrafish have an embryonic to larval *globin* switch

In all vertebrates studied, which include humans and mice, a complex process known as hemoglobin switching has been shown to exist (Bauer and Orkin, 2011; McConnell et al., 2011). The ISHs presented in this thesis, along with qPCR data from the Zon laboratory, indicates that this process also exists in the zebrafish (Ganis et al., unpublished) (Figure 4). The ISHs, completed through 5dpf, confirm that there is an embryonic to larval *globin* switch. Through the use of ISH, important features of this switch are displayed, particularly the decrease in embryonic *globins*, such as β_{e3}, and the increase in larval ones, like β_{e2}.

This ISH data allows for a fuller characterization of the hemoglobin switching process in the zebrafish. With the addition of this information, there is a better understanding of this complex process in the zebrafish and there can be further use of the zebrafish as a model of hemoglobin switching.

4.1.1 Advantages of ISH

The ISH technique provides important information about the nature of *globin* expression that qPCR alone is unable to detect. ISH presents both spatial and temporal expression patterns for the *globins*. This is particularly important for the hematopoietic system, since the location of blood cells is a significant feature of their development (Orkin and Zon, 2008). At the 25 ss and 24 hpf stages, the ISH highlights that hematopoiesis occurs in the ICM at these stages (Figure 4). Since circulation does not

begin until soon after 24 hpf, the *globin* staining seen before 24 hpf also indicates where red blood cell production is occurring, although *globins*, especially once circulation begins, do not generally stain for the location of blood production (Orkin and Zon, 2008). The ISH data also displays a level of sensitivity that qPCR does not provide. In the ISHs for both α_{a1} and β_{a1}, particularly evident at 24 hpf, a few cells are stained indicating that the zebrafish embryos are expressing the respective adult *globins* (Figure 4). This feature of ISH, high resolution at the single-cell level, provides information that qPCR is unable. With so few cells expressing these adult *globins*, qPCR would not detect these levels of expression or the data would be considered unreliable. Additionally, ISH can distinguish between results that qPCR would be unable to differentiate because it visualizes expressing cells. One result indicating that a few cells are strongly expressing versus another result demonstrating that many cells are weakly expressing can seem the same according to qPCR, since it only calculates the total amount of the cDNA in question from the entire embryo. ISH has the ability to clarify these results and display two distinct expression patterns.

4.1.2 Limitations of ISH

While ISH is a useful technique for high resolution tissue staining, there are some limitations to its usage. Since one of the important factors in the success of an ISH is the ability of the RNA probe to penetrate the tissue, the capacity to use whole-mount ISH as the zebrafish embryo ages decreases. Although protinase K is used to allow the penetration of the *globin* RNA probes, when the embryo becomes too old, at about 7-8 dpf for the zebrafish, protinase K is no longer effective. This limited the use of whole-

mount ISH to demonstrating only the embryonic to larval *globin* switch, since this switch occurs before 7-8 dpf. It is possible to use ISH after this time point, but to ensure accuracy and a complete expression pattern, it must be done on histological sections (Thisse and Thisse, 2008). Although characterization of the larval to adult switch through ISH can be done using histological sections, with nine *globins* as well as a range of time points between 22 dpf (toward the end of the zebrafish larval stages) and 32 dpf (when the adult *globin* expression patterns have been nearly completely established) necessary for a full characterization (Ganis et al., unpublished)(Figures 2B and 2C), it was not feasible to complete within the scope of this thesis.

An additional limitation of ISH is its usefulness for discerning differences in staining intensity as the gene becomes more strongly expressed. While ISH does well to display differences in gene expression patterns at low to moderate stain intensities, it becomes difficult to distinguish differences at higher levels of expression. When considering the increases in α_{e3} and β_{e3} expression due to *SOX6* overexpression, for example, it was difficult to determine which embryos displayed increased expression versus those that did not, particularly since the increased expression pattern looked somewhat like overstaining or non-specific staining (Figure 7). Similarly, when examining decreases in α_{e3} and β_{e3} expression following Sox6 knockdown, it was challenging to discern between embryos with morphological changes but without a decrease in stain intensity and embryos with both features (Figure 7). As a result of this ambiguity, and the stringency necessary to prevent falsely declaring results significant when they are not, it is possible that the scoring of embryos with expression changes versus those without is inaccurate. In situations with a small sample size or an even more

ambiguous expression pattern, results could be determined insignificant when they may actually be important.

The differences in probes and how they affect ISH results, especially among the *globins*, highlight another limitation of the use ISH to compare among various probes. The amount of time necessary for staining to be complete for the *globin* probes ranged from 20 minutes to 3 hours, indicating how sensitive some of the probes where. Additionally, some probes stained extremely well, such as β_{e3}, so that they could be reused very often, while others demonstrated dramatic decreases in efficiency of staining within a small number of uses. With the results of ISHs being so sensitive to the probe used, ISH is much more useful to compare results within the same probe versus among different probes. These limitations exemplify why the combination of qPCR and ISH is ideal for a full characterization of globin expression.

ISH also displays the variability of gene expression patterns found among zebrafish clutches, which can be both a hindrance and an advantage. Through the repetition of ISHs for the same *globins* and same developmental time points while generating Figure 4, I noticed that there were small differences in staining intensity that could be accounted for due to variability among different clutches. Regardless, this variability did not have a significant effect on these results, since the changes in expression that confirm *globin* switching were more dramatic than the relatively small changes in expression seen due to variability. The advantage of demonstrating this variability through ISH is noted when significant results are found on the small scale. In a screen, for example, if there were a 50 percent knockdown of a gene in only 25 percent of the embryos seen, the knockdown may not be noticed in qPCR since all of the cDNA is

pooled together. With an ISH, however, embryos could be examined on a single embryo basis to isolate the important hits.

4.2 Bcl11a and Sox6 regulate *globins* in the zebrafish embryo

Multiple studies indicate that BCL11A and SOX6 have a critical role in the silencing of embryonic *globins* in adult human and murine cells (Dumitriu et al., 2006; Sankaran et al., 2008; Sankaran et al., 2009; Wilber et al., 2011; Xu et al., 2010; Yi et al., 2006). Until this thesis, however, the roles of Bcl11a and Sox6 as regulators of *globin* gene expression in the embryo were unexplored. The data discussed in 3.2 and 3.3 indicate that both of these transcription factors regulate *globins* in the zebrafish embryo.

4.2.1 Prior knowledge on the regulation of *globin* genes

A mechanism involving a combination of autonomous gene control and gene competition for direct interaction with the LCR has been suggested as one method for the regulation of hemoglobin switching (Palstra et al., 2008). The idea of autonomous gene control is based on studies that, through the use human *globin* genes with only proximal *cis*-regulatory elements in transgenic mice, demonstrated that the human transgenes express with the correct developmental timing in a position-dependent manner (Palstra et al., 2008). Gene competition was first suggested in patients with HPFH who saw an increase in γ-gene expression due to mutations in the γ-*globin* genes. Along with this, patients saw a decrease in β-*globin* gene expression from the same mutated allele, but normal expression in the nonmutated allele (Giglioni et al., 1984). This competition for LCR function was confirmed with experiments indicating that gene order and relative

distance were relevant. Introduction of a β-gene between the LCR and γ-gene causes early activation of the β-gene and decreased expression of the γ-gene. Additionally, introduction of a second marked β-gene in the locus leads to higher expression of the more proximal gene with no increase in the total expression of the two β-*globins* together (Noordermeer and de Laat, 2008; Palstra et al., 2008). For both humans and mice, the LCR upregulates only one gene at a time, with the genes competing with each other for activation (Noordermeer and de Laat, 2008). With the identification of a functional LCR, the zebrafish is likely to regulate its *globin* genes through the LCR in a similar manner (Ganis et al., unpublished).

4.2.2 Regulation of adult *globins*

The ISH data presented in 3.1, as well as unpublished data from Jared Ganis in the Zon laboratory, show that a small number of adult *globin* expressing cells are found in the zebrafish embryo (Figure 4). Additionally, knockdown and overexpression of Bcl11a in the zebrafish embryo led to decreases and increases, respectively, in α_{a1} expression (Ganis, unpublished data). The data presented in 3.2.1 confirm the data from the Zon laboratory (Figure 5). I completed similar experiments for Sox6 revealing that overexpression of *SOX6* in the zebrafish embryo resulted in an increase in α_{a1} expressing cells (Figure 7). This data suggests that Bcl11a and Sox6 regulate adult *globin* expression in the zebrafish embryo.

These results fall in line with what is known thus far concerning the regulation of *globin* switching. In adult human cells, BCL11A occupies several discrete regions within the human β-*globin* locus, including the LCR, allowing for the reconfiguration of the

human β-*globin* cluster (Sankaran et al., 2008; Xu et al., 2010). In the presence of BCL11A, the LCR acts preferentially with the adult *globin* gene, while the absence of BCL11A leads to the LCR favoring the fetal genes (Xu et al., 2010). Additionally, since BCL11A may silence γ-genes though SOX6 (Xu et al., 2010), for both BCL11A and SOX6 interaction through the LCR, direct and indirect, is important for the selection of which *globin* genes will be expressed. Under the gene competition mechanism, it is not surprising that knockdown of Bcl11a leads to a decrease in adult *globins* or that overexpression of *BCL11A* and *SOX6* causes an increase in adult *globins*. Knockdown of Bcl11a would cause the LCR to favor embryonic genes, thus causing diminished transcription of the adult *globins*. Similarly, overexpression of either *BCL11A* or *SOX6* would cause increased transcription of the adult *globins* due to preferential interaction of the LCR with the adult globin genes. These results support this model of regulation. Considering the regulation of embryonic *globins*, however, as will be discussed in the following section, this model of regulation may not be accurate.

4.2.3 Regulation of embryonic *globins*

Surprisingly, the results presented in 3.2.2 and 3.3.2 indicate that knockdown of Bcl11a or Sox6 leads to a decrease in embryonic *globins*, while overexpression of either causes an increase in embryonic *globins* (Figures 6 and 7). These results mirror the changes seen in adult *globin* expression with knockdown and overexpression of Bcl11a or Sox6 in the zebrafish embryo. Considering the prior knowledge of the gene competition model and the interactions of BCL11A and SOX6 in the human adult cells, it is unexpected that both embryonic and adult *globin* expression would change in the same

manner upon manipulation of BCL11A or SOX6. This change in regulation from what we see in the adult mouse and human cells could be the result of a different mechanism in the zebrafish as compared to mammals or a different mechanism in embryos as compared to adults.

4.2.4 Possible mechanisms of regulation

Although the zebrafish has two *globin* switches like humans, there are differences between the human and zebrafish *globin* clusters that may point to this difference in regulation. Unlike humans with α-like and β-like loci arranged 5' to 3' in the order of expression during development (McConnell et al., 2011)(Figure 1A), the zebrafish *globin* genes are located in major and minor loci and are arranged in α/β pairs oriented in a 5'-3' to 3'-5' manner (Ganis et al., unpublished)(Figure 2A). This change in *globin* gene arrangement may require a different regulatory mechanism unlike that seen in mammals, causing adult and embryonic *globin* expression to change in a similar manner upon manipulation of regulatory factors. This locus difference being main reason for the results presented in this thesis seems unlikely, however, since regardless of orientation and arrangement in α/β pairs, the *globin* genes are still clustered together according to when the *globins* are expressed during development (Figure 2A). Therefore, the LCR favoring one *globin* cluster over the other should still result in the favored cluster having increased expression, while the other cluster would have decreased expression.

It is possible that the gene competition model in which the presence of a transcription factor causes the LCR to preferentially bind a certain gene is not applicable to the zebrafish. In mice, nascent RNA signals from either the embryonic *globins* or adult

globins are often detected simultaneously at one allele, which the authors attribute to rapid switching of the LCR between genes on the same allele (Trimborn et al., 1999). A similar mechanism could be occurring in the zebrafish, in which the LCR regulates more than one gene at a time, switching equally between the adult and embryonic *globin* genes. Knockdown or overexpression of the factors mediating the LCR binding could lead to less or more effective binding, respectively, of the LCR to both the adult and embryonic *globin* genes. This would result in the same changes in expression for both adult and embryonic *globin* clusters like the data seen in this thesis.

This manner of regulation for Bcl11a and Sox6 could also be the result of differences in embryonic versus adult regulation. One possibility is that Bcl11a and Sox6 do not function in the same manner in the embryo as they do in the adult. Differing expression of BCL11A isoforms in human adult versus embryonic cells suggest that this may be the case. While adult human cells expressed the long forms of BCL11A (L and XL), human embryonic cells expressed very little, if any, of the long forms. Instead, they expressed shorter variants (Sankaran et al., 2008). Zebrafish may also express multiple variants of Bcl11a that regulate globin expression differently between adult and embryonic *globins*. A ZFIN search indicates that Bcl11a is found in two forms: Bcl11aa and Bcl11ab. All experiments were done considering Bcl11aa, leaving Bcl11ab as an unexplored variant that may have a regulatory role in the zebrafish embryo.

It could be possible that Bcl11ab regulates zebrafish embryonic *globins* while Bcl11aa regulates adult. There is also the possibility that alternate splicing occurs for either or both of these Bcl11a variants, and any number of these forms could be regulating embryonic *globins*, while the alternate versions regulate adult. The differences

in these variants may not have been detected, however, because the experiments done in this thesis may have been affecting multiple variants of Bcl11a without my knowledge. The Bcl11aa MO, for example, is an Exon 1 MO that could be affecting multiple variants. In humans, all variants of BCL11A have exons 1 and 2 in common (Liu et al., 2006). In order to knockdown a specific variant, later exons would have to be targeted. In the zebrafish, Bcl11aa could be arranged similar to human BCL11A, causing this exon MO to affect all alternatively spliced versions of Bcl11aa. Additionally, to explore the option that Bcl11ab is involved, a MO targeting Bcl11ab specifically would have to be used to see if it caused changes in embryonic *globins* alone, suggesting that Bcl11ab is restricted to the zebrafish embryo.

Similarly, overexpression experiments with endogenous zebrafish Bcl11a and human BCL11A-L and –XL could have the issue of affecting a different variant of Bcl11a than expected. Zebrafish and human overexpession resulted in the same expression patterns, but they may all be long versions of Bcl11a that lead to increases in both embryonic and adult *globins*. A short variant, both the zebrafish and adult versions, may only affect the embryonic *globins*. Additionally, since Bcl11a may be operating through Sox6 regardless of which form it is in, any manipulations of Sox6 could affect all variants of Bcl11a.

4.2.5 Future Experiments

Experiments to elucidate the relationship between Bcl11a and Sox6, as well as clarify the roles of different variants of Bcl11a in the zebrafish embryo, are necessary to more fully understand hemoglobin switching in the zebrafish embryo. A recent study

indicates that BCL11A and SOX6 physically interact and knockdown of BCL11A and SOX6 together in adult primary human erythroblasts leads to greater HbF production than seen with either knockdown alone (Xu et al., 2011). Although knockdown of Bcl11a and Sox6 alone were completed in this thesis, knockdown of the two together was not done. Injections of a mix of Bcl11a MO and Sox6 MO should be done to explore the possibility that these transcription factors cooperate in the zebrafish embryo. This relationship could also be made clearer through knockdown of one factor while overexpressing another. Knockdown of Bcl11a and overexpression of *SOX6* simultaneously, then knockdown of Sox6 while overexpressing *Bcl11a* will provide information about the necessity of either transcription factor for effective regulation of the *globins*. I made two attempts to overexpress zebrafish *Bcl11a* and knockdown Sox6, but every injected embryo died before 24 hpf in both attempts. High concentration doses can cause embryo death; the highest concentration (6ng) of Sox6 MO resulted in the death of all injected embryos for every attempt. Therefore, alterations in the concentrations of the expression vector and MO may need to be made to prevent toxicity causing such a high rate of embryo death

Acquiring qPCR data confirming all changes in expression pattern observed, as well as providing quantitative information about how much the expression of each *globin* changed would be an important experiment to complete. As discussed earlier in 4.1.2, limitations of ISH include its decreased ability to discern differences in staining intensity as the gene becomes more strongly expressed as well as its diminished usefulness when comparing expression patterns between clutches. qPCR data would be an important confirmation of increases or decreases in *globin* expression in those globins that have

high wildtype expression levels at 24 hpf. α_{e3} and β_{e3}, for example, are strongly expressed at 24 hpf, so I had to be very conservative when deciding that an injected embryo was demonstrating an increase or decrease in expression (Figures 6 and 7). qPCR would remove any of the ambiguity involved with determining if there was truly an increase or decrease in expression and account for the limitation of ISH concerning its ability to distinguish between varying levels of relatively strong expression. Additionally, overexpression of human *BCL11A-XL* seemed particularly effective as indicated by its darker β_{e2} staining as compared to any other overexpression experiment (Figure 6). While this may be an accurate result, it could also be due to the variability among clutches; the uninjected expression of β_{e2} for that clutch seemed slightly darker than any other uninjected embryos also stained for β_{e2}. qPCR data would be able to confirm if overexpression of human *BCL11A-XL* actually produces more β_{e2} expressing cells than other overexpression experiments by quantifying the increase as compared to uninjected siblings. This quantification could be easily compared among all *BCL11A* overexpression experiments.

To determine if Bcl11ab, an alternate version of Bcl11aa, may be involved with the regulation of embryonic *globins*, experiments with a MO and expression vector targeting Bcl11ab specifically should be carried out. A Bcl11ab MO was obtained and I attempted to see if there were differences in *globin* expression between Bcl11aa and Bcl11ab, but was unable to get conclusive results before the submission of this thesis. Expression vectors driving *Bcl11ab* expression as well as the short variants of human *BCL11A* should also be used to examine the differences, if any, among embryonic and adult *globin* expression as compared to the alternate variants used in this thesis. The

results from these experiments could suggest different methods of *globin* regulation in the adult versus embryo dependent on distinct BCL11A variants for both humans and zebrafish.

All of these proposed experiments would contribute to the data presented in this thesis as well as help determine the regulatory mechanism behind *globin* switching in the zebrafish embryo.

5. Conclusion and Future Directions

The results presented in this thesis confirm, through the use of ISH, that zebrafish undergo the complex process of *globin* switching, specifically the embryonic to larval *globin* switch. This thesis also indicates that Bcl11a and Sox6 regulate *globin* expression in the zebrafish embryo. Knockdown of Bcl11a and Sox6 leads to a decrease in both adult and embryonic *globin* expression, while overexpression of both factors causes an increase in adult and embryonic *globin* expression. These results suggest that the manner of *globin* regulation is different from that seen in adult human and murine cells, since knockdown of either BCL11A and SOX6 in adult human and murine cells leads to an increase of embryonic *globins*, the reverse of that demonstrated in the zebrafish embryo. This difference in *globin* regulation may be attributed to regulatory differences in the zebrafish or in the embryo. Future experiments with different variants of both zebrafish and human BCL11A will help narrow down where the main difference lies.

Consideration should also be given to a third transcription factor that, more recently, has been implicated as an important target to increase HbF in vivo: KLF1. Sankaran *et al.* refer to the "KFL1-BCL11A-SOX6 HbF silencing pathway," highlighting the importance of KLF1 by mentioning that, in humans, in vivo heterozygous disruption, due to mutations, of KLF1, but not in SOX6, leads to HPFH (Sankaran et al., 2011). Exploring all three of these transcription factors together may lead to a more thorough understanding of hemoglobin switching as well as determine the most effective therapeutic target to ameliorate hemoglobinopathies.

REFERENCES CITED

Bauer, D.E., and Orkin, S.H. (2011). Update on fetal hemoglobin gene regulation in hemoglobinopathies. Curr Opin Pediatr *23*, 1-8.

Brownlie, A., Hersey, C., Oates, A.C., Paw, B.H., Falick, A.M., Witkowska, H.E., Flint, J., Higgs, D., Jessen, J., Bahary, N., Zhu, H., Lin, S., and Zon, L.I. (2003). Characterization of embryonic globin genes of the zebrafish. Dev. Biol. *255*, 48-61.

Cohen-Barak, O., Erickson, D.T., Badowski, M.S., Fuchs, D.A., Klassen, C.L., Harris, D.T., and Brilliant, M.H. (2007). Stem cell transplantation demonstrates that Sox6 represses epsilon y globin expression in definitive erythropoiesis of adult mice. Exp Hematol *35*, 358-367.

Dumitriu, B., Patrick, M.R., Petschek, J.P., Cherukuri, S., Klingmuller, U., Fox, P.L., and Lefebvre, V. (2006). Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. Blood *108*, 1198-1207.

Giglioni, B., Casini, C., Mantovani, R., Merli, S., Comi, P., Ottolenghi, S., Saglio, G., Camaschella, C., and Mazza, U. (1984). A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta-thalassemia. Embo J *3*, 2641-2645.

Hsia, N., and Zon, L.I. (2005). Transcriptional regulation of hematopoietic stem cell development in zebrafish. Exp Hematol *33*, 1007-1014.

Jing, L., and Zon, L.I. (2011). Zebrafish as a model for normal and malignant hematopoiesis. Dis Model Mech *4*, 433-438.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev Dyn *203*, 253-310.

Kwan, K.M., Fujimoto, E., Grabher, C., Mangum, B.D., Hardy, M.E., Campbell, D.S., Parant, J.M., Yost, H.J., Kanki, J.P., and Chien, C.B. (2007). The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. Dev Dyn *236*, 3088-3099.

Leikin, S.L., Gallagher, D., Kinney, T.R., Sloane, D., Klug, P., and Rida, W. (1989). Mortality in children and adolescents with sickle cell disease. Cooperative Study of Sickle Cell Disease. Pediatrics *84*, 500-508.

Liu, H., Ippolito, G.C., Wall, J.K., Niu, T., Probst, L., Lee, B.S., Pulford, K., Banham, A.H., Stockwin, L., Shaffer, A.L., *et al.* (2006). Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. Mol Cancer *5*, 18.

McConnell, S.C., Huo, Y., Liu, S., and Ryan, T.M. (2011). Human globin knock-in mice complete fetal-to-adult hemoglobin switching in postnatal development. Mol Cell Biol *31*, 876-883.

Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene 'knockdown' in zebrafish. Nat Genet *26*, 216-220.

Noordermeer, D., and de Laat, W. (2008). Joining the loops: beta-globin gene regulation. IUBMB Life *60*, 824-833.

Orkin, S.H., and Higgs, D.R. (2010). Medicine. Sickle cell disease at 100 years. Science *329*, 291-292.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Paik, E.J., and Zon, L.I. (2010). Hematopoietic development in the zebrafish. Int J Dev Biol *54*, 1127-1137.

Palis, J. and Segel, G.B. (1998). Developmental biology of erythropoiesis. Blood Rev. *12,* 106-114.

Palstra, R.J., de Laat, W., and Grosveld, F. (2008). Beta-globin regulation and long-range interactions. Adv Genet *61*, 107-142.

Rombough, P., and Drader, H. (2009). Hemoglobin enhances oxygen uptake in larval zebrafish (Danio rerio) but only under conditions of extreme hypoxia. J Exp Biol *212*, 778-784.

Sankaran, V.G. (2011). Targeted therapeutic strategies for fetal hemoglobin induction. Hematology Am Soc Hematol Educ Program *2011*, 459-465.

Sankaran, V.G., Menne, J., and Heller, R. (2011). Heterozygous disruption of human SOX6 is insufficient to impair erythropoiesis or silencing of fetal hemoglobin. Blood *117*, 4396-4397.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839-1842.

Sankaran, V.G., Xu, J., and Orkin, S.H. (2010). Transcriptional silencing of fetal hemoglobin by BCL11A. Ann. N.Y. Acad. Sci. *1202,* 64-68.

Sankaran, V.G., Xu, J., Ragoczy, T., Ippolito, G.C., Walkley, C.R., Maika, S.D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M.A., *et al.* (2009). Developmental and species-divergent globin switching are driven by BCL11A. Nature *460*, 1093-1097.

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. Exp Hematol *33*, 259-271.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat Protoc *3*, 59-69.

Trimborn, T., Gribnau, J., Grosveld, F., and Fraser, P. (1999). Mechanisms of developmental control of transcription in the murine alpha- and beta-globin loci. Genes Dev *13*, 112-124.

Wilber, A., Hargrove, P.W., Kim, Y.S., Riberdy, J.M., Sankaran, V.G., Papanikolaou, E., Georgomanoli, M., Anagnou, N.P., Orkin, S.H., Nienhuis, A.W., *et al.* (2011). Therapeutic levels of fetal hemoglobin in erythroid progeny of beta-thalassemic CD34+ cells after lentiviral vector-mediated gene transfer. Blood *117*, 2817-2826.

Wilber, A., Nienhuis, A.W., and Persons DA. (2011). Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. Blood. *117,* 3945-3953.

Xu, J., Peng, C., Sankaran, V.G., Shao, Z., Esrick, E.B., Chong, B.G., Ippolito, G.C., Fujiwara, Y., Ebert, B.L., Tucker, P.W., *et al.* (2011). Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. Science *334*, 993-996.

Xu, J., Sankaran, V.G., Ni, M., Menne, T.F., Puram, R.V., Kim, W., and Orkin, S.H. (2010). Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev *24*, 783-798.

Yi, Z., Cohen-Barak, O., Hagiwara, N., Kingsley, P.D., Fuchs, D.A., Erickson, D.T., Epner, E.M., Palis, J., and Brilliant, M.H. (2006). Sox6 directly silences epsilon globin expression in definitive erythropoiesis. PLoS Genet *2*, e14.

Zhou, D., Liu, K., Sun, C.W., Pawlik, K.M., and Townes, T.M. (2010). KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat Genet *42*, 742-744.